

SEQUENCE LISTING

5 (1) GENERAL INFORMATION:

(i) APPLICANT: Bell, Graeme I.
Yamagata, Kazuya
Oda, Naohisha
Kaisaki, Pamela J.
Furuta, Hiroto
Horikawa, Yukio
Menzel, Stephen

10 (ii) TITLE OF INVENTION: MUTATIONS IN THE DIABETES SUSCEPTIBILITY
GENES HEPATOCYTE NUCLEAR FACTOR (HNF) 1 ALPHA, HNF-1BETA
AND HNF-4ALPHA

15 (iii) NUMBER OF SEQUENCES: 146

20 (iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Arnold, White & Durkee
(B) STREET: P.O. Box 4433
(C) CITY: Houston
(D) STATE: Texas
(E) COUNTRY: USA
(F) ZIP: 77210

25 (v) COMPUTER READABLE FORM:

30 (A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

35 (vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: US Unknown
(B) FILING DATE: Concurrently Herewith
(C) CLASSIFICATION: Unknown

40 (vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 60/029,679
(B) FILING DATE: 30-OCT-1996

45 (vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 60/028,056
(B) FILING DATE: 02-OCT-1996

50 (vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 60/025,719
(B) FILING DATE: 10-SEP-1996

55 (viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Wilson, Mark B.
(B) REGISTRATION NUMBER: 37,259
(C) REFERENCE/DOCKET NUMBER: ARCD:272

60 (ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 512/418-3000
(B) TELEFAX: 512/474-7577

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 3238 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

10 (ix) FEATURE:

- 10 (A) NAME/KEY: modified_base..
(B) LOCATION: 988
(D) OTHER INFORMATION: /mod_base= OTHER

/note= "N = A, C, G, or T"

15 (ix) FEATURE:

- 15 (A) NAME/KEY: CDS
(B) LOCATION: join(24..986, 990..1916)

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CGTGGCCCTG	TGGCAGCCGA	GCC ATG GTT TCT AAA CTG AGC CAG CTG CAG	50
		Met Val Ser Lys Leu Ser Gln Leu Gln	
		1 5	
ACG GAG CTC CTG GCG GCC CTG CTC GAG TCA GGG CTG AGC AAA GAG GCA	98		
Thr Glu Leu Leu Ala Ala Leu Leu Glu Ser Gly Leu Ser Lys Glu Ala			
10 15 20 25			
CTG ATC CAG GCA CTG GGT GAG CCG GGG CCC TAC CTC CTG GCT GGA GAA	146		
Leu Ile Gln Ala Leu Gly Glu Pro Gly Pro Tyr Leu Leu Ala Gly Glu			
30 35 40			
GGC CCC CTG GAC AAG GGG GAG TCC TGC GGC GGC GGT CGA GGG GAG CTG	194		
Gly Pro Leu Asp Lys Gly Glu Ser Cys Gly Gly Arg Gly Glu Leu			
45 50 55			
GCT GAG CTG CCC AAT GGG CTG GGG GAG ACT CGG GGC TCC GAG GAC GAG	242		
Ala Glu Leu Pro Asn Gly Leu Gly Glu Thr Arg Gly Ser Glu Asp Glu			
60 65 70			
ACG GAC GAC GAT GGG GAA GAC TTC ACG CCA CCC ATC CTC AAA GAG CTG	290		
Thr Asp Asp Asp Gly Glu Asp Phe Thr Pro Pro Ile Leu Lys Glu Leu			
75 80 85			
GAG AAC CTC AGC CCT GAG GAG GCG GCC CAC CAG AAA GCC GTG GTG GAG	338		
Glu Asn Leu Ser Pro Glu Glu Ala Ala His Gln Lys Ala Val Val Glu			
90 95 100 105			
ACC CTT CTG CAG GAG GAC CCG TGG CGT GTG GCG AAG ATG GTC AAG TCC	386		
Thr Leu Leu Gln Glu Asp Pro Trp Arg Val Ala Lys Met Val Lys Ser			
110 115 120			
TAC CTG CAG CAG CAC AAC ATC CCA CAG CGG GAG GTG GTC GAT ACC ACT	434		
Tyr Leu Gln Gln His Asn Ile Pro Gln Arg Glu Val Val Asp Thr Thr			
125 130 135			
GGC CTC AAC CAG TCC CAC CTG TCC CAA CAC CTC AAC AAG GGC ACT CCC	482		
Gly Leu Asn Gln Ser His Leu Ser Gln His Leu Asn Lys Gly Thr Pro			
140 145 150			
ATG AAG ACG CAG AAG CGG GCC CTG TAC ACC TGG TAC GTC CGC AAG	530		

	Met Lys Thr Gln Lys Arg Ala Ala Leu Tyr Thr Trp Tyr Val Arg Lys		
	155 160 165		
5	CAG CGA GAG GTG GCG CAG CAG TTC ACC CAT GCA GGG CAG GGA GGG CTG Gln Arg Glu Val Ala Gln Gln Phe Thr His Ala Gly Gln Gly Gly Leu 170 175 180 185		578
10	ATT GAA GAG CCC ACA GGT GAT GAG CTA CCA ACC AAG AAG GGG CGG AGG Ile Glu Glu Pro Thr Gly Asp Glu Leu Pro Thr Lys Lys Gly Arg Arg 190 195 200		626
15	AAC CGT TTC AAG TGG GGC CCA GCA TCC CAG CAG ATC CTG TTC CAG GCC Asn Arg Phe Lys Trp Gly Pro Ala Ser Gln Gln Ile Leu Phe Gln Ala 205 210 215		674
20	TAT GAG AGG CAG AAG AAC CCT AGC AAG GAG GAG CGA GAG ACG CTA GTG Tyr Glu Arg Gln Lys Asn Pro Ser Lys Glu Glu Arg Glu Thr Leu Val 220 225 230		722
25	GAG GAG TGC AAT AGG GCG GAA TGC ATC CAG AGA GGG GTG TCC CCA TCA Glu Glu Cys Asn Arg Ala Glu Cys Ile Gln Arg Gly Val Ser Pro Ser 235 240 245		770
30	CAG GCA CAG GGG CTG GGC TCC AAC CTC GTC ACG GAG GTG CGT GTC TAC Gln Ala Gln Gly Leu Gly Ser Asn Leu Val Thr Glu Val Arg Val Tyr 250 255 260 265		818
35	AAC TGG TTT GCC AAC CGG CGC AAA GAA GAA GCC TTC CGG CAC AAG CTG Asn Trp Phe Ala Asn Arg Arg Lys Glu Glu Ala Phe Arg His Lys Leu 270 275 280		866
40	GCC ATG GAC ACG TAC AGC GGG CCC CCC CCA GGG CCA GGC CCG GGA CCT Ala Met Asp Thr Tyr Ser Gly Pro Pro Pro Gly Pro Gly Pro Gly Pro 285 290 295		914
45	GCG CTG CCC GCT CAC AGC TCC CCT GGC CTG CCT CCA CCT GCC CTC TCC Ala Leu Pro Ala His Ser Ser Pro Gly Leu Pro Pro Pro Ala Leu Ser 300 305 310		962
50	CCC AGT AAG GTC CAC GGT GTG CGC TNT GGA CAG CCT GCG ACC AGT GAG Pro Ser Lys Val His Gly Val Arg Gly Gln Pro Ala Thr Ser Glu 315 320 325		1010
55	ACT GCA GAA GTA CCC TCA AGC AGC GGC GGT CCC TTA GTG ACA GTG TCT Thr Ala Glu Val Pro Ser Ser Ser Gly Gly Pro Leu Val Thr Val Ser 330 335 340		1058
60	ACA CCC CTC CAC CAA GTG TCC CCC ACG GGC CTG GAG CCC AGC CAC AGC Thr Pro Leu His Gln Val Ser Pro Thr Gly Leu Glu Pro Ser His Ser 345 350 355 360		1106
	CTG CTG AGT ACA GAA AAG CTG GTC TCA GCA GCT GGG GGC CCC CTC Leu Leu Ser Thr Glu Ala Lys Leu Val Ser Ala Ala Gly Gly Pro Leu 365 370 375		1154
	CCC CCT GTC AGC ACC CTG ACA GCA CTG CAC AGC TTG GAG CAG ACA TCC Pro Pro Val Ser Thr Leu Thr Ala Leu His Ser Leu Glu Gln Thr Ser 380 385 390		1202
	CCA GGC CTC AAC CAG CAG CCC CAG AAC CTC ATC ATG GCC TCA CTT CCT Pro Gly Leu Asn Gln Gln Pro Gln Asn Leu Ile Met Ala Ser Leu Pro		1250

	395	400	405	
5	GGG GTC ATG ACC ATC GGG CCT GGT GAG CCT GCC TCC CTG GGT CCT ACG Gly Val Met Thr Ile Gly Pro Gly Glu Pro Ala Ser Leu Gly Pro Thr 410	415	420	1298
10	TTC ACC AAC ACA GGT GCC TCC ACC CTG GTC ATC GGC CTG GCC TCC ACG Phe Thr Asn Thr Gly Ala Ser Thr Leu Val Ile Gly Leu Ala Ser Thr 425	430	435	1346
15	CAG GCA CAG AGT GTG CCG GTC ATC AAC AGC ATG GGC AGC AGC CTG ACC Gln Ala Gln Ser Val Pro Val Ile Asn Ser Met Gly Ser Ser Leu Thr 445	450	455	1394
20	ACC CTG CAG CCC GTC CAG TTC TCC CAG CCG CTG CAC CCC TCC TAC CAG Thr Leu Gln Pro Val Gln Phe Ser Gln Pro Leu His Pro Ser Tyr Gln 460	465	470	1442
25	CAG CCG CTC ATG CCA CCT GTG CAG AGC CAT GTG ACC CAG AGC CCC TTC Gln Pro Leu Met Pro Pro Val Gln Ser His Val Thr Gln Ser Pro Phe 475	480	485	1490
30	ATG GCC ACC ATG GCT CAG CTG CAG AGC CCC CAC GCC CTC TAC AGC CAC Met Ala Thr Met Ala Gln Leu Gln Ser Pro His Ala Leu Tyr Ser His 490	495	500	1538
35	AAG CCC GAG GTG GCC CAG TAC ACC CAC ACG GGC CTG CTC CCG CAG ACT Lys Pro Glu Val Ala Gln Tyr Thr His Thr Gly Leu Leu Pro Gln Thr 505	510	515	1586
40	ATG CTC ATC ACC GAC ACC AAC CTG AGC GCC CTG GCC AGC CTC ACG Met Leu Ile Thr Asp Thr Thr Asn Leu Ser Ala Leu Ala Ser Leu Thr 525	530	535	1634
45	CCC ACC AAG CAG GTC TTC ACC TCA GAC ACT GAG GCC TCC AGT GAG TCC Pro Thr Lys Gln Val Phe Thr Ser Asp Thr Glu Ala Ser Ser Glu Ser 540	545	550	1682
50	GGG CTT CAC ACG CCG GCA TCT CAG GCC ACC ACC CTC CAC GTC CCC AGC Gly Leu His Thr Pro Ala Ser Gln Ala Thr Thr Leu His Val Pro Ser 555	560	565	1730
55	CAG GAC CCT GCC GGC ATC CAG CAC CTG CAG CCG GCC CAC CGG CTC AGC Gln Asp Pro Ala Gly Ile Gln His Leu Gln Pro Ala His Arg Leu Ser 570	575	580	1778
60	GCC AGC CCC ACA GTG TCC TCC AGC AGC CTG GTG CTG TAC CAG AGC TCA Ala Ser Pro Thr Val Ser Ser Ser Leu Val Leu Tyr Gln Ser Ser 585	590	595	1826
	GAC TCC AGC AAT GGC CAG AGC CAC CTG CTG CCA TCC AAC CAC AGC GTC Asp Ser Ser Asn Gly Gln Ser His Leu Leu Pro Ser Asn His Ser Val 605	610	615	1874
	ATC GAG ACC TTC ATC TCC ACC CAG ATG GCC TCT TCC TCC CAG Ile Glu Thr Phe Ile Ser Thr Gln Met Ala Ser Ser Gln 620	625	630	1916
	TAACCACGGC ACCTGGGCC TGCGCTGT ACTGCCTGCT TGGGGGTGA TGAGGGCAGC AGCCAGCCCT GCCTGGAGGA CCTGAGCCTG CCGAGCAACC GTGGCCCTTC CTGGACAGCT			1976
				2036

	GTGCCTCGCT	CCCCACTCTG	CTCTGATGCA	TCAGAAAGGG	AGGGCTCTGA	GGCGCCCAA	2096
5	CCCGTGGAGG	CTGCTCGGGG	TGCACAGGAG	GGGGTCGTGG	AGAGCTAGGA	GCAAAGCCTG	2156
	TTCATGGCAG	ATGTAGGAGG	GACTGTCGCT	GCTTCGTGGG	ATACAGTCTT	CTTACTTGGA	2216
10	ACTGAAGGGG	GC GG C CT ATG	ACTTGGGCAC	CCCCAGCCTG	GGCCTATGGA	GAGCCCTGGG	2276
	ACCGCTACAC	CACTCTGGCA	GCCACACTTC	TCAGGACACA	GGCCTGTGTA	GCTGTGACCT	2336
15	GCTGAGCTCT	GAGAGGCCCT	GGATCAGCGT	GGCCTTGTTC	TGTCACCAAT	GTACCCACCG	2396
	GGCCACTCCT	TCCTGCCCA	ACTCCTTCCA	GCTAGTGACC	CACATGCCAT	TTGTACTGAC	2456
20	CCCATCACCT	ACTCACACAG	GCATTTCTG	GGTGGCTACT	CTGTGCCAGA	GCCTGGGCT	2516
	CTAACTGCCT	GAGCCCAGGG	AGGCCGAAGC	TAACAGGGAA	GGCAGGCAGG	GCTCTCCTGG	2576
25	TCTTCCCATC	CCCAGCGATT	CCCTCTCCCA	GGCCCCATGA	CCTCCAGCTT	TCCTGTATTT	2636
	CTTCCCAGA	GCATGATGCC	TCTGAGGCCA	GCCTGGCCTC	CTGCCTCTAC	TGGGAAGGCT	2696
30	ACTTCGGGGC	TGGGAAGTCG	TCCTTACTCC	TGTGGGAGCC	TCGCAACCCG	TGCCAAGTCC	2756
	AGGTCCCTGGT	GGGGCAGCTC	CTCTGTCTCG	AGCGCCCTGC	AGACCCTGCC	CTTGTGTTGGG	2816
35	GCAGGAGTAG	CTGAGTCAC	AAGGCAGCAA	GGCCCGAGCA	GCTGAGCAGG	GCCGGGGAAC	2876
	TGGCCAAGCT	GAGGTGCCA	GGAGAAGAAA	GAGGTGACCC	CAGGGCACAG	GAGCTACCTG	2936
40	TGTGGACAGG	ACTAACACTC	AGAAGCCTGG	GTGCCTGGCT	GGCTGAGGGC	AGTCGCAGC	2996
	CACCCCTGAGG	AGTCTGAGGT	CCTGAGCACT	GCCAGGAGGG	ACAAAGGAGC	CTGTGAACCC	3056
45	AGGACAAGCA	TGGTCCCACA	TCCCTGGGCC	TGCTGCTGAG	AACCTGGCCT	TCAGTGTACC	3116
	GCGTCTACCC	TGGGATTCAAG	GAAAAGGCCT	GGGGTGACCC	GGCACCCCT	GCAGCTTGTA	3176
50	GCCAGCCGGG	GCGAGTGGCA	CGTTTATTCA	ACTTTAGTA	AAGTCAAGGA	GAAATGCGGT	3236
	GG					-3238	

45 (2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 630 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Val Ser Lys Leu Ser Gln Leu Gln Thr Glu Leu Leu Ala Ala Leu
 1 5 10 15

60 Leu Glu Ser Gly Leu Ser Lys Glu Ala Leu Ile Gln Ala Leu Gly Glu
 20 25 30

Pro Gly Pro Tyr Leu Leu Ala Gly Glu Gly Pro Leu Asp Lys Gly Glu
 35 40 45
 5 Ser Cys Gly Gly Arg Gly Glu Leu Ala Glu Leu Pro Asn Gly Leu
 50 55 60
 Gly Glu Thr Arg Gly Ser Glu Asp Glu Thr Asp Asp Asp Gly Glu Asp
 65 70 75 80
 10 Phe Thr Pro Pro Ile Leu Lys Glu Leu Glu Asn Leu Ser Pro Glu Glu
 85 90 95
 Ala Ala His Gln Lys Ala Val Val Glu Thr Leu Leu Gln Glu Asp Pro
 15 100 105 110
 Trp Arg Val Ala Lys Met Val Lys Ser Tyr Leu Gln Gln His Asn Ile
 115 120 125
 20 Pro Gln Arg Glu Val Val Asp Thr Thr Gly Leu Asn Gln Ser His Leu
 130 135 140
 Ser Gln His Leu Asn Lys Gly Thr Pro Met Lys Thr Gln Lys Arg Ala
 25 145 150 155 160
 Ala Leu Tyr Thr Trp Tyr Val Arg Lys Gln Arg Glu Val Ala Gln Gln
 165 170 175
 Phe Thr His Ala Gly Gln Gly Leu Ile Glu Glu Pro Thr Gly Asp
 180 185 190
 30 Glu Leu Pro Thr Lys Lys Gly Arg Arg Asn Arg Phe Lys Trp Gly Pro
 195 200 205
 Ala Ser Gln Gln Ile Leu Phe Gln Ala Tyr Glu Arg Gln Lys Asn Pro
 35 210 215 220
 Ser Lys Glu Glu Arg Glu Thr Leu Val Glu Glu Cys Asn Arg Ala Glu
 225 230 235 240
 40 Cys Ile Gln Arg Gly Val Ser Pro Ser Gln Ala Gln Gly Leu Gly Ser
 245 250 255
 Asn Leu Val Thr Glu Val Arg Val Tyr Asn Trp Phe Ala Asn Arg Arg
 260 265 270
 45 Lys Glu Glu Ala Phe Arg His Lys Leu Ala Met Asp Thr Tyr Ser Gly
 275 280 285
 Pro Pro Pro Gly Pro Gly Pro Gly Pro Ala Leu Pro Ala His Ser Ser
 50 290 295 300
 Pro Gly Leu Pro Pro Pro Ala Leu Ser Pro Ser Lys Val His Gly Val
 305 310 315 320
 55 Arg Gly Gln Pro Ala Thr Ser Glu Thr Ala Glu Val Pro Ser Ser Ser
 325 330 335
 Gly Gly Pro Leu Val Thr Val Ser Thr Pro Leu His Gln Val Ser Pro
 340 345 350
 60 Thr Gly Leu Glu Pro Ser His Ser Leu Leu Ser Thr Glu Ala Lys Leu

	355	360	365	
	Val Ser Ala Ala Gly Gly Pro Leu Pro Pro Val Ser Thr Leu Thr Ala			
5	370	375	380	
	Leu His Ser Leu Glu Gln Thr Ser Pro Gly Leu Asn Gln Gln Pro Gln			
	385	390	395	400
10	Asn Leu Ile Met Ala Ser Leu Pro Gly Val Met Thr Ile Gly Pro Gly			
	405	410	415	
	Glu Pro Ala Ser Leu Gly Pro Thr Phe Thr Asn Thr Gly Ala Ser Thr			
	420	425	430	
15	Leu Val Ile Gly Leu Ala Ser Thr Gln Ala Gln Ser Val Pro Val Ile			
	435	440	445	
	Asn Ser Met Gly Ser Ser Leu Thr Thr Leu Gln Pro Val Gln Phe Ser			
20	450	455	460	
	Gln Pro Leu His Pro Ser Tyr Gln Gln Pro Leu Met Pro Pro Val Gln			
	465	470	475	480
25	Ser His Val Thr Gln Ser Pro Phe Met Ala Thr Met Ala Gln Leu Gln			
	485	490	495	
	Ser Pro His Ala Leu Tyr Ser His Lys Pro Glu Val Ala Gln Tyr Thr			
30	500	505	510	
	His Thr Gly Leu Leu Pro Gln Thr Met Leu Ile Thr Asp Thr Thr Asn			
	515	520	525	
35	Leu Ser Ala Leu Ala Ser Leu Thr Pro Thr Lys Gln Val Phe Thr Ser			
	530	535	540	
	Asp Thr Glu Ala Ser Ser Glu Ser Gly Leu His Thr Pro Ala Ser Gln			
40	545	550	555	560
	Ala Thr Thr Leu His Val Pro Ser Gln Asp Pro Ala Gly Ile Gln His			
	565	570	575	
	Leu Gln Pro Ala His Arg Leu Ser Ala Ser Pro Thr Val Ser Ser Ser			
	580	585	590	
45	Ser Leu Val Leu Tyr Gln Ser Ser Asp Ser Ser Asn Gly Gln Ser His			
	595	600	605	
	Leu Leu Pro Ser Asn His Ser Val Ile Glu Thr Phe Ile Ser Thr Gln			
50	610	615	620	
	Met Ala Ser Ser Ser Gln			
	625	630		

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3238 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

10 (ix) FEATURE:

- (A) NAME/KEY: modified_base
(B) LOCATION: 988
(D) OTHER INFORMATION: /mod_base= OTHER

/note= "N = A, C, G, or T"

15 (ix) FEATURE:

- (A) NAME/KEY: CDS
(B) LOCATION: join(24..986, 990..1916)

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CGTGGCCCTG	TGGCAGCCGA	GCC	ATG	GTT	TCT	AAA	CTG	AGC	CAG	CTG	CAG	50				
		Met	Val	Ser	Lys	Leu	Ser	Gln	Leu	Gln						
		1				5										
ACG	GAG	CTC	CTG	GCG	GCC	CTG	CTC	GAG	TCA	GGG	CTG	AGC	AAA	GAG	GCA	98
Thr	Glu	Leu	Leu	Ala	Ala	Leu	Leu	Glu	Ser	Gly	Leu	Ser	Lys	Glu	Ala	
10		15				20							25			
CTG	ATC	CAG	GCA	CTG	GGT	GAG	CCG	GGG	CCC	TAC	CTC	CTG	GCT	GGA	GAA	146
Leu	Ile	Gln	Ala	Leu	Gly	Glu	Pro	Gly	Pro	Tyr	Leu	Leu	Ala	Gly	Glu	
		30			35								40			
GGC	CCC	CTG	GAC	AAG	GGG	GAG	TCC	TGC	GGC	GGC	GGT	CGA	GGG	GAG	CTG	194
Gly	Pro	Leu	Asp	Lys	Gly	Glu	Ser	Cys	Gly	Gly	Gly	Arg	Gly	Glu	Leu	
		45			50							55				
GCT	GAG	CTG	CCC	AAT	GGG	CTG	GGG	GAG	ACT	CGG	GGC	TCC	GAG	GAC	GAG	242
Ala	Glu	Leu	Pro	Asn	Gly	Leu	Gly	Glu	Thr	Arg	Gly	Ser	Glu	Asp	Glu	
		60			65							70				
ACG	GAC	GAC	GAT	GGG	GAA	GAC	TTC	ACG	CCA	CCC	ATC	CTC	AAA	GAG	CTG	290
Thr	Asp	Asp	Asp	Gly	Glu	Asp	Phe	Thr	Pro	Pro	Ile	Leu	Lys	Glu	Leu	
		75			80							85				
GAG	AAC	CTC	AGC	CCT	GAG	GCG	GCC	CAC	CAG	AAA	GCC	GTG	GTG	GAG	338	
Glu	Asn	Leu	Ser	Pro	Glu	Glu	Ala	Ala	His	Gln	Lys	Ala	Val	Val	Glu	
		90			95					100			105			
ACC	CTT	CTG	CAG	GAG	GAC	CCG	TGG	CGT	GTG	GCG	AAG	ATG	GTC	AAG	TCC	386
Thr	Leu	Leu	Gln	Glu	Asp	Pro	Trp	Arg	Val	Ala	Lys	Met	Val	Lys	Ser	
		110			115							120				
TAC	CTG	CAG	CAG	CAC	AAC	ATC	CCA	CAG	CAG	GAG	GTG	GTC	GAT	ACC	ACT	434
Tyr	Leu	Gln	Gln	His	Asn	Ile	Pro	Gln	Gln	Glu	Val	Val	Asp	Thr	Thr	
		125			130							135				
GGC	CTC	AAC	CAG	TCC	CAC	CTG	TCC	CAA	CAC	CTC	AAC	AAG	GGC	ACT	CCC	482
Gly	Leu	Asn	Gln	Ser	His	Leu	Ser	Gln	His	Leu	Asn	Lys	Gly	Thr	Pro	
		140			145							150				

	ATG AAG ACG CAG AAG CGG GCC GCC CTG TAC ACC TGG TAC GTC CGC AAG Met Lys Thr Gln Lys Arg Ala Ala Leu Tyr Thr Trp Tyr Val Arg Lys 155 160 165	530
5	CAG CGA GAG GTG GCG CAG CAG TTC ACC CAT GCA GGG CAG GGA GGG CTG Gln Arg Glu Val Ala Gln Gln Phe Thr His Ala Gly Gln Gly Gly Leu 170 175 180 185	578
10	ATT GAA GAG CCC ACA GGT GAT GAG CTA CCA ACC AAG AAG GGG CGG AGG Ile Glu Glu Pro Thr Gly Asp Glu Leu Pro Thr Lys Lys Gly Arg Arg 190 195 200	626
15	AAC CGT TTC AAG TGG GGC CCA GCA TCC CAG CAG ATC CTG TTC CAG GCC Asn Arg Phe Lys Trp Gly Pro Ala Ser Gln Gln Ile Leu Phe Gln Ala 205 210 215	674
20	TAT GAG AGG CAG AAG AAC CCT AGC AAG GAG GAG CGA GAG ACG CTA GTG Tyr Glu Arg Gln Lys Asn Pro Ser Lys Glu Glu Arg Glu Thr Leu Val 220 225 230	722
25	GAG GAG TGC AAT AGG GCG GAA TGC ATC CAG AGA GGG GTG TCC CCA TCA Glu Glu Cys Asn Arg Ala Glu Cys Ile Gln Arg Gly Val Ser Pro Ser 235 240 245	770
30	CAG GCA CAG GGG CTG GGC TCC AAC CTC GTC ACG GAG GTG CGT GTC TAC Gln Ala Gln Gly Leu Gly Ser Asn Leu Val Thr Glu Val Arg Val Tyr 250 255 260 265	818
35	AAC TGG TTT GCC AAC CGG CGC AAA GAA GAA GCC TTC CGG CAC AAG CTG Asn Trp Phe Ala Asn Arg Arg Lys Glu Glu Ala Phe Arg His Lys Leu 270 275 280	866
40	GCC ATG GAC ACG TAC AGC GGG CCC CCC CCA GGG CCA GGC CCG GGA CCT Ala Met Asp Thr Tyr Ser Gly Pro Pro Pro Gly Pro Gly Pro Gly Pro 285 290 295	914
45	GCG CTG CCC GCT CAC AGC TCC CCT GGC CTG CCT CCA CCT GCC CTC TCC Ala Leu Pro Ala His Ser Ser Pro Gly Leu Pro Pro Ala Leu Ser 300 305 310	962
50	CCC AGT AAG GTC CAC GGT GTG CGC TNT GGA CAG CCT GCG ACC AGT GAG Pro Ser Lys Val His Gly Val Arg Gly Gln Pro Ala Thr Ser Glu 315 320 325	1010
55	ACT GCA GAA GTA CCC TCA AGC AGC GGC GGT CCC TTA GTG ACA GTG TCT Thr Ala Glu Val Pro Ser Ser Ser Gly Gly Pro Leu Val Thr Val Ser 330 335 340	1058
60	ACA CCC CTC CAC CAA GTG TCC CCC ACG GGC CTG GAG CCC AGC CAC AGC Thr Pro Leu His Gln Val Ser Pro Thr Gly Leu Glu Pro Ser His Ser 345 350 355 360	1106
	CTG CTG AGT ACA GAA GCC AAG CTG GTC TCA GCA GCT GGG GGC CCC CTC Leu Leu Ser Thr Glu Ala Lys Leu Val Ser Ala Ala Gly Gly Pro Leu 365 370 375	1154
	CCC CCT GTC AGC ACC CTG ACA GCA CTG CAC AGC TTG GAG CAG ACA TCC Pro Pro Val Ser Thr Leu Thr Ala Leu His Ser Leu Glu Gln Thr Ser 380 385 390	1202

	CCA GGC CTC AAC CAG CAG CCC CAG AAC CTC ATC ATG GCC TCA CTT CCT Pro Gly Leu Asn Gln Gln Pro Gln Asn Leu Ile Met Ala Ser Leu Pro 395 400 405	1250
5	GGG GTC ATG ACC ATC GGG CCT GGT GAG CCT GCC TCC CTG GGT CCT ACG Gly Val Met Thr Ile Gly Pro Gly Glu Pro Ala Ser Leu Gly Pro Thr 410 415 420	1298
10	TTC ACC AAC ACA GGT GCC TCC ACC CTG GTC ATC GGC CTG GCC TCC ACG Phe Thr Asn Thr Gly Ala Ser Thr Leu Val Ile Gly Leu Ala Ser Thr 425 430 435 440	1346
15	CAG GCA CAG AGT GTG CCG GTC ATC AAC AGC ATG GGC AGC AGC CTG ACC Gln Ala Gln Ser Val Pro Val Ile Asn Ser Met Gly Ser Ser Leu Thr 445 450 455	1394
20	ACC CTG CAG CCC GTC CAG TTC TCC CAG CCG CTG CAC CCC TCC TAC CAG Thr Leu Gln Pro Val Gln Phe Ser Gln Pro Leu His Pro Ser Tyr Gln 460 465 470	1442
25	CAG CCG CTC ATG CCA CCT GTG CAG AGC CAT GTG ACC CAG AGC CCC TTC Gln Pro Leu Met Pro Pro Val Gln Ser His Val Thr Gln Ser Pro Phe 475 480 485	1490
30	ATG GCC ACC ATG GCT CAG CTG CAG AGC CCC CAC GCC CTC TAC AGC CAC Met Ala Thr Met Ala Gln Leu Gln Ser Pro His Ala Leu Tyr Ser His 490 495 500	1538
35	AAG CCC GAG GTG GCC CAG TAC ACC CAC ACG GGC CTG CTC CCG CAG ACT Lys Pro Glu Val Ala Gln Tyr Thr His Thr Gly Leu Leu Pro Gln Thr 505 510 515 520	1586
40	ATG CTC ATC ACC GAC ACC AAC CTG AGC GCC CTG GCC AGC CTC ACG Met Leu Ile Thr Asp Thr Thr Asn Leu Ser Ala Leu Ala Ser Leu Thr 525 530 535	1634
45	CCC ACC AAG CAG GTC TTC ACC TCA GAC ACT GAG GCC TCC AGT GAG TCC Pro Thr Lys Gln Val Phe Thr Ser Asp Thr Glu Ala Ser Ser Glu Ser 540 545 550	1682
50	GGG CTT CAC ACG CCG GCA TCT CAG GCC ACC ACC CTC CAC GTC CCC AGC Gly Leu His Thr Pro Ala Ser Gln Ala Thr Thr Leu His Val Pro Ser 555 560 565	1730
55	CAG GAC CCT GCC GGC ATC CAG CAC CTG CAG CCG GCC CAC CGG CTC AGC Gln Asp Pro Ala Gly Ile Gln His Leu Gln Pro Ala His Arg Leu Ser 570 575 580	1778
60	GCC AGC CCC ACA GTG TCC TCC AGC AGC CTG GTG CTG TAC CAG AGC TCA Ala Ser Pro Thr Val Ser Ser Ser Leu Val Leu Tyr Gln Ser Ser 585 590 595 600	1826
65	GAC TCC AGC AAT GGC CAG AGC CAC CTG CTG CCA TCC AAC CAC AGC GTC Asp Ser Ser Asn Gly Gln Ser His Leu Leu Pro Ser Asn His Ser Val 605 610 615	1874
70	ATC GAG ACC TTC ATC TCC ACC CAG ATG GCC TCT TCC CAG Ile Glu Thr Phe Ile Ser Thr Gln Met Ala Ser Ser Ser Gln 620 625 630	1916
75	TAACCACGGC ACCTGGGCC TGGGGCCTGT ACTGCCTGCT TGGGGGGTGA TGAGGGCAGC	1976

	AGCCAGCCCT	GCCTGGAGGA	CCTGAGCCTG	CCGAGCAACC	GTGGCCCTTC	CTGGACAGCT	2036
5	GTGCCTCGCT	CCCCACTCTG	CTCTGATGCA	TCAGAAAGGG	AGGGCTCTGA	GGCGCCCCAA	2096
	CCCGTGGAGG	CTGCTCGGGG	TGCACAGGAG	GGGGTCGTGG	AGAGCTAGGA	GCAAAGCCTG	2156
	TTCATGGCAG	ATGTAGGAGG	GACTGTCGCT	GCTTCGTGGG	ATACAGTCTT	CTTACTTGGA	2216
10	ACTGAAGGGG	GCGGCCTATG	ACTTGGGCAC	CCCCAGCCTG	GGCCTATGGA	GAGCCCTGGG	2276
	ACCGCTACAC	CACTCTGGCA	GCCACACTTC	TCAGGACACA	GGCCTGTGTA	GCTGTGACCT	2336
15	GCTGAGCTCT	GAGAGGCCCT	GGATCAGCGT	GGCCTTGTTC	TGTCACCAAT	GTACCCACCG	2396
	GGCCACTCCT	TCCTGCCCCA	ACTCCTTCCA	GCTAGTGACC	CACATGCCAT	TTGTACTGAC	2456
	CCCATCACCT	ACTCACACAG	GCATTTCTG	GGTGGCTACT	CTGTGCCAGA	GCCTGGGGCT	2516
20	CTAACTGCCT	GAGCCCAGGG	AGGCCGAAGC	TAACAGGGAA	GGCAGGCAGG	GCTCTCCTGG	2576
	TCTTCCCATC	CCCAGCGATT	CCCTCTCCC	GGCCCCATGA	CCTCCAGCTT	TCCTGTATTT	2636
25	CTTCCAAGA	GCATGATGCC	TCTGAGGCCA	GCCTGGCCTC	CTGCCTCTAC	TGGGAAGGCT	2696
	ACTTCGGGGC	TGGGAAGTCG	TCCTTACTCC	TGTGGGAGCC	TCGCAACCCG	TGCCAAGTCC	2756
	AGGTCCCTGGT	GGGCAGCTC	CTCTGTCTCG	AGCGCCCTGC	AGACCCTGCC	CTTGTGTTGGG	2816
30	GCAGGAGTAG	CTGAGCTCAC	AAGGCAGCAA	GGCCCGAGCA	GCTGAGCAGG	GCCGGGAAAC	2876
	TGGCCAAGCT	GAGGTGCCA	GGAGAAGAAA	GAGGTGACCC	CAGGGCACAG	GAGCTACCTG	2936
35	TGTGGACAGG	ACTAACACTC	AGAAGCCTGG	GTGCCTGGCT	GGCTGAGGGC	AGTTCGCAGC	2996
	CACCCCTGAGG	AGTCTGAGGT	CCTGAGCACT	GCCAGGAGGG	ACAAAGGAGC	CTGTGAACCC	3056
	AGGACAAGCA	TGGTCCCACA	TCCCTGGGCC	TGCTGCTGAG	AACCTGGCCT	TCAGTGTACC	3116
40	GCGTCTACCC	TGGGATTCAAG	GAAAAGGCCT	GGGGTGACCC	GGCACCCCT	GCAGCTTGTA	3176
	GCCAGCCGGG	GCGAGTGGCA	CGTTTATTAA	ACTTTTAGTA	AAGTCAAGGA	GAAATGCGGT	3236
45	GG						3238

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 630 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
 - ii) MOLECULE TYPE: protein
 - xii) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met	Val	Ser	Lys	Leu	Ser	Gln	Leu	Gln	Thr	Glu	Leu	Leu	Ala	Ala	Leu
1				5					10					15	
Leu	Glu	Ser	Gly	Leu	Ser	Lys	Glu	Ala	Leu	Ile	Gln	Ala	Leu	Gly	Glu

	20	25	30
5	Pro Gly Pro Tyr Leu Leu Ala Gly Glu Gly Pro Leu Asp Lys Gly Glu 35	40	45
10	Ser Cys Gly Gly Gly Arg Gly Glu Leu Ala Glu Leu Pro Asn Gly Leu 50	55	60
15	Gly Glu Thr Arg Gly Ser Glu Asp Glu Thr Asp Asp Asp Gly Glu Asp 65	70	75
20	Phe Thr Pro Pro Ile Leu Lys Glu Leu Glu Asn Leu Ser Pro Glu Glu 85	90	95
25	Ala Ala His Gln Lys Ala Val Val Glu Thr Leu Leu Gln Glu Asp Pro 100	105	110
30	Trp Arg Val Ala Lys Met Val Lys Ser Tyr Leu Gln Gln His Asn Ile 115	120	125
35	Pro Gln Gln Glu Val Val Asp Thr Thr Gly Leu Asn Gln Ser His Leu 130	135	140
40	Ser Gln His Leu Asn Lys Gly Thr Pro Met Lys Thr Gln Lys Arg Ala 145	150	155
45	Ala Leu Tyr Thr Trp Tyr Val Arg Lys Gln Arg Glu Val Ala Gln Gln 165	170	175
50	Phe Thr His Ala Gly Gln Gly Gly Leu Ile Glu Glu Pro Thr Gly Asp 180	185	190
55	Glu Leu Pro Thr Lys Lys Gly Arg Arg Asn Arg Phe Lys Trp Gly Pro 195	200	205
60	Ala Ser Gln Gln Ile Leu Phe Gln Ala Tyr Glu Arg Gln Lys Asn Pro 210	215	220
65	Ser Lys Glu Glu Arg Glu Thr Leu Val Glu Glu Cys Asn Arg Ala Glu 225	230	235
70	Cys Ile Gln Arg Gly Val Ser Pro Ser Gln Ala Gln Gly Leu Gly Ser 245	250	255
75	Asn Leu Val Thr Glu Val Arg Val Tyr Asn Trp Phe Ala Asn Arg Arg 260	265	270
80	Lys Glu Glu Ala Phe Arg His Lys Leu Ala Met Asp Thr Tyr Ser Gly 275	280	285
85	Pro Pro Pro Gly Pro Gly Pro Gly Pro Ala Leu Pro Ala His Ser Ser 290	295	300
90	Pro Gly Leu Pro Pro Pro Ala Leu Ser Pro Ser Lys Val His Gly Val 305	310	315
95	Arg Gly Gln Pro Ala Thr Ser Glu Thr Ala Glu Val Pro Ser Ser Ser 325	330	335
100	Gly Gly Pro Leu Val Thr Val Ser Thr Pro Leu His Gln Val Ser Pro 340	345	350

Thr Gly Leu Glu Pro Ser His Ser Leu Leu Ser Thr Glu Ala Lys Leu
355 360 365

5 Val Ser Ala Ala Gly Gly Pro Leu Pro Pro Val Ser Thr Leu Thr Ala
370 375 380

Leu His Ser Leu Glu Gln Thr Ser Pro Gly Leu Asn Gln Gln Pro Gln
385 390 395 400

10 Asn Leu Ile Met Ala Ser Leu Pro Gly Val Met Thr Ile Gly Pro Gly
405 410 415

15 Glu Pro Ala Ser Leu Gly Pro Thr Phe Thr Asn Thr Gly Ala Ser Thr
420 425 430

Leu Val Ile Gly Leu Ala Ser Thr Gln Ala Gln Ser Val Pro Val Ile
435 440 445

20 Asn Ser Met Gly Ser Ser Leu Thr Thr Leu Gln Pro Val Gln Phe Ser
450 455 460

Gln Pro Leu His Pro Ser Tyr Gln Gln Pro Leu Met Pro Pro Val Gln
465 470 475 480

25 Ser His Val Thr Gln Ser Pro Phe Met Ala Thr Met Ala Gln Leu Gln
485 490 495

30 Ser Pro His Ala Leu Tyr Ser His Lys Pro Glu Val Ala Gln Tyr Thr
500 505 510

His Thr Gly Leu Leu Pro Gln Thr Met Leu Ile Thr Asp Thr Thr Asn
515 520 525

35 Leu Ser Ala Leu Ala Ser Leu Thr Pro Thr Lys Gln Val Phe Thr Ser
530 535 540

40 Asp Thr Glu Ala Ser Ser Glu Ser Gly Leu His Thr Pro Ala Ser Gln
545 550 555 560

Ala Thr Thr Leu His Val Pro Ser Gln Asp Pro Ala Gly Ile Gln His
565 570 575

45 Leu Gln Pro Ala His Arg Leu Ser Ala Ser Pro Thr Val Ser Ser Ser
580 585 590

Ser Leu Val Leu Tyr Gln Ser Ser Asp Ser Ser Asn Gly Gln Ser His
595 600 605

50 Leu Leu Pro Ser Asn His Ser Val Ile Glu Thr Phe Ile Ser Thr Gln
610 615 620

Met Ala Ser Ser Ser Gln
625 630

55

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3239 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

10 (ix) FEATURE:

- (A) NAME/KEY: modified_base
(B) LOCATION: 989
(D) OTHER INFORMATION: /mod_base= OTHER

/note= "N = A, C, G, or T"

15 (ix) FEATURE:

- (A) NAME/KEY: CDS
(B) LOCATION: 24..965

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CGTGGCCCTG	TGGCAGCCGA	GCC	ATG	GTT	TCT	AAA	CTG	AGC	CAG	CTG	CAG	50										
							Met	Val	Ser	Lys	Leu	Ser	Gln	Leu	Gln							
							1			5												
ACG	GAG	CTC	CTG	GCG	GCC	CTG	CTC	GAG	TCA	GGG	CTG	AGC	AAA	GAG	GCA							
															98							
							Thr	Glu	Leu	Leu	Ala	Ala	Leu	Leu	Glu	Ser	Gly	Leu	Ser	Lys	Glu	Ala
							10		15		20		25									
CTG	ATC	CAG	GCA	CTG	GGT	GAG	CCG	GGG	CCC	TAC	CTC	CTG	GCT	GGA	GAA	146						
							Leu	Ile	Gln	Ala	Leu	Gly	Glu	Pro	Gly	Pro	Tyr	Leu	Leu	Ala	Gly	Glu
							30		35		40											
GGC	CCC	CTG	GAC	AAG	GGG	GAG	TCC	TGC	GGC	GGC	GGT	CGA	GGG	GAG	CTG	194						
							Gly	Pro	Leu	Asp	Lys	Gly	Glu	Ser	Cys	Gly	Gly	Arg	Gly	Glu	Leu	
							45		50		55											
GCT	GAG	CTG	CCC	AAT	GGG	CTG	GGG	GAG	ACT	CGG	GGC	TCC	GAG	GAC	GAG	242						
							Ala	Glu	Leu	Pro	Asn	Gly	Leu	Gly	Glu	Thr	Arg	Gly	Ser	Glu	Asp	Glu
							60		65		70											
ACG	GAC	GAC	GAT	GGG	GAA	GAC	TTC	ACG	CCA	CCC	ATC	CTC	AAA	GAG	CTG	290						
							Thr	Asp	Asp	Asp	Gly	Glu	Asp	Phe	Thr	Pro	Pro	Ile	Leu	Lys	Glu	Leu
							75		80		85											
GAG	AAC	CTC	AGC	CCT	GAG	GAG	GCG	GCC	CAC	CAG	AAA	GCC	GTG	GTG	GAG	338						
							Glu	Asn	Leu	Ser	Pro	Glu	Ala	Ala	His	Gln	Lys	Ala	Val	Val	Glu	
							90		95		100		105									
ACC	CTT	CTG	CAG	GAG	GAC	CCG	TGG	CGT	GTG	GCG	AAG	ATG	GTC	AAG	TCC	386						
							Thr	Leu	Gln	Glu	Asp	Pro	Trp	Arg	Val	Ala	Lys	Met	Val	Lys	Ser	
							110		115		120											
TAC	CTG	CAG	CAG	CAC	AAC	ATC	CCA	CAG	CGG	GAG	GTG	GTC	GAT	ACC	ACT	434						
							Tyr	Leu	Gln	Gln	His	Asn	Ile	Pro	Gln	Arg	Glu	Val	Val	Asp	Thr	Thr
							125		130		135											
GGC	CTC	AAC	CAG	TCC	CAC	CTG	TCC	CAA	CAC	CTC	AAC	AAG	GGC	ACT	CCC	482						
							Gly	Leu	Asn	Gln	Ser	His	Leu	Ser	Gln	His	Leu	Asn	Lys	Gly	Thr	Pro
							140		145		150											

	ATG AAG ACG CAG AAG CGG GCC GCC CTG TAC ACC TGG TAC GTC CGC AAG Met Lys Thr Gln Lys Arg Ala Ala Leu Tyr Thr Trp Tyr Val Arg Lys 155 160 165	530
5	CAG CGA GAG GTG GCG CAG CAG TTC ACC CAT GCA GGG CAG GGA GGG CTG Gln Arg Glu Val Ala Gln Phe Thr His Ala Gly Gln Gly Gly Leu 170 175 180 185	578
10	ATT GAA GAG CCC ACA GGT GAT GAG CTA CCA ACC AAG AAG GGG CGG AGG Ile Glu Glu Pro Thr Gly Asp Glu Leu Pro Thr Lys Lys Gly Arg Arg 190 195 200	626
15	AAC CGT TTC AAG TGG GGC CCA GCA TCC CAG CAG ATC CTG TTC CAG GCC Asn Arg Phe Lys Trp Gly Pro Ala Ser Gln Gln Ile Leu Phe Gln Ala 205 210 215	674
20	TAT GAG AGG CAG AAG AAC CCT AGC AAG GAG GAG CGA GAG ACG CTA GTG Tyr Glu Arg Gln Lys Asn Pro Ser Lys Glu Glu Arg Glu Thr Leu Val 220 225 230	722
25	GAG GAG TGC AAT AGG GCG GAA TGC ATC CAG AGA GGG GTG TCC CCA TCA Glu Glu Cys Asn Arg Ala Glu Cys Ile Gln Arg Gly Val Ser Pro Ser 235 240 245	770
30	CAG GCA CAG GGG CTG GGC TCC AAC CTC GTC ACG GAG GTG CGT GTC TAC Gln Ala Gln Gly Leu Gly Ser Asn Leu Val Thr Glu Val Arg Val Tyr 250 255 260 265	818
35	AAC TGG TTT GCC AAC CGG CGC AAA GAA GAA GCC TTC CGG CAC AAG CTG Asn Trp Phe Ala Asn Arg Arg Lys Glu Ala Phe Arg His Lys Leu 270 275 280	866
40	GCC ATG GAC ACG TAC AGC GGG CCC CCC AGG GCC AGG CCC GGG ACC Ala Met Asp Thr Tyr Ser Gly Pro Pro Pro Arg Ala Arg Pro Gly Thr 285 290 295	914
45	TGC GCT GCC CGC TCA CAG CTC CCC TGG CCT GCC TCC ACC TGC CCT CTC Cys Ala Ala Arg Ser Gln Leu Pro Trp Pro Ala Ser Thr Cys Pro Leu 300 305 310	962
50	CCC CAGTAAGGTC CACGGTGTGC GCTNTGGACA GCCTGCGACC AGTGAGACTG Pro	1015
55	CAGAAGTACC CTCAAGCAGC GGCGGTCCCT TAGTGACAGT GTCTACACCC CTCCACCAAG TGTCCCCCAC GGGCCTGGAG CCCAGCCACA GCCTGCTGAG TACAGAAGCC AAGCTGGTCT	1075 1135
60	CAGCAGCTGG GGGCCCCCTC CCCCCCTGTCA GCACCCGTAC AGCACTGCAC AGCTTGGAGC AGACATCCCC AGGCCTCAAC CAGCAGCCCC AGAACCTCAT CATGGCCTCA CTTCCCTGGGG TCATGACCAT CGGGCCTGGT GAGCCTGCCT CCCTGGGTCC TACGTTCACC AACACAGGTG CCTCCACCCCT GGTCATCGGC CTGGCCTCCA CGCAGGCACA GAGTGTGCCG GTCATCAACA GCATGGGCAG CAGCCTGACC ACCCTGCAGC CCGTCCAGTT CTCCCAGCCG CTGCACCCCT CCTACCAGCA GCCGCTCATG CCACCTGTGC AGAGCCATGT GACCCAGAGC CCCTTCATGG	1195 1255 1315 1375 1435 1495

	CCACCATGGC TCAGCTGCAG AGCCCCACG CCCTCTACAG CCACAAGCCC GAGGTGGCCC	1555
	AGTACACCCA CACGGGCCTG CTCCCGCAGA CTATGCTCAT CACCGACACC ACCAACCTGA	1615
5	GCGCCCTGGC CAGCCTCACG CCCACCAAGC AGGTCTTCAC CTCAGACACT GAGGCCTCCA	1675
	GTGAGTCCGG GCTTCACACG CCGGCATCTC AGGCCACCAC CCTCCACGTC CCCAGCCAGG	1735
10	ACCCCTGCCGG CATCCAGCAC CTGCAGCCGG CCCACCGGCT CAGCGCCAGC CCCACAGTGT	1795
	CCTCCAGCAG CCTGGTGCTG TACCAGAGCT CAGACTCCAG CAATGGCCAG AGCCACCTGC	1855
	TGCCATCCAA CCACAGCGTC ATCGAGACCT TCATCTCCAC CCAGATGGCC TCTTCCTCCC	1915
15	AGTAACCACG GCACCTGGC CCTGGGGCCT GTACTGCCTG CTTGGGGGT GATGAGGGCA	1975
	GCAGCCAGCC CTGCCTGGAG GACCTGAGCC TGCGAGCAA CCGTGGCCCT TCCTGGACAG	2035
20	CTGTGCCTCG CTCCCCACTC TGCTCTGATG CATCAGAAAG GGAGGGCTCT GAGGCGCCCC	2095
	AACCCGTGGA GGCTGCTCGG GGTGCACAGG AGGGGGTCGT GGAGAGCTAG GAGCAAAGCC	2155
	TGTTCATGGC AGATGTAGGA GGGACTGTCTG CTGCTTCGTG GGATAACAGTC TTCTTACTTG	2215
25	GAACACTGAAGG GGGCGGCCTA TGACTTGGC ACCCCCAGCC TGGGCCTATG GAGAGCCCTG	2275
	GGACCGCTAC ACCACTCTGG CAGCCACACT TCTCAGGACA CAGGCCTGTG TAGCTGTGAC	2335
	CTGCTGAGCT CTGAGAGGCC CTGGATCAGC GTGGCCTTGT TCTGTCACCA ATGTACCCAC	2395
30	CGGGCCACTC CTTCCCTGCC CAACTCCTTC CAGCTAGTGA CCCACATGCC ATTTGTACTG	2455
	ACCCCATCAC CTACTCACAC AGGCATTCC TGGGTGGCTA CTCTGTGCCA GAGCCTGGGG	2515
35	CTCTAACTGC CTGAGCCCAG GGAGGCCGAA GCTAACAGGG AAGGCAGGCA GGGCTCTCCT	2575
	GGTCTTCCA TCCCCAGCGA TTCCCTCTCC CAGGCCCAT GACCTCCAGC TTTCCTGTAT	2635
	TTCTTCCA GAGCATGATG CCTCTGAGGC CAGCCTGGCC TCCTGCCTCT ACTGGGAAGG	2695
40	CTACTTCGGG GCTGGGAAGT CGTCCTTA CTCAGCTGGAG CCTCGCAACC CGTGCCAAGT	2755
	CCAGGTCCCTG GTGGGGCAGC TCCTCTGTCT CGAGGCCCT GCAGACCCTG CCCTTGTGG	2815
45	GGGCAGGAGT AGCTGAGCTC ACAAGGCAGC AAGGCCGAG CAGCTGAGCA GGGCCGGGGA	2875
	ACTGGCCAAG CTGAGGTGCC CAGGAGAAGA AAGAGGTGAC CCCAGGGCAC AGGAGCTACC	2935
	TGTGTGGACA GGACTAACAC TCAGAACGCT GGGTGCCTGG CTGGCTGAGG GCAGTCGCA	2995
50	GCCACCCCTGA GGAGTCTGAG GTCCTGAGCA CTGCCAGGAG GGACAAAGGA GCCTGTGAAC	3055
	CCAGGACAAG CATGGTCCA CATCCCTGGG CCTGCTGCTG AGAACCTGGC CTTCAGTGT	3115
55	CCCGGTCTAC CCTGGGATTC AGGAAAAGGC CTGGGGTGAC CCGGCACCCCC CTGCAGCTTG	3175
	TAGCCAGCCG GGGCGAGTGG CACGTTTATT TAAACTTTAG TAAAGTCAAG GAGAAATGCG	3235
	GTGA	3239

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 314 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: protein

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met Val Ser Lys Leu Ser Gln Leu Gln Thr Glu Leu Leu Ala Ala Leu
1 5 10 15

15 Leu Glu Ser Gly Leu Ser Lys Glu Ala Leu Ile Gln Ala Leu Gly Glu
20 25 30

Pro Gly Pro Tyr Leu Leu Ala Gly Glu Gly Pro Leu Asp Lys Gly Glu
35 40 45

20 Ser Cys Gly Gly Arg Gly Glu Leu Ala Glu Leu Pro Asn Gly Leu
50 55 60

25 Gly Glu Thr Arg Gly Ser Glu Asp Glu Thr Asp Asp Asp Gly Glu Asp
65 70 75 80

Phe Thr Pro Pro Ile Leu Lys Glu Leu Glu Asn Leu Ser Pro Glu Glu
85 90 95

30 Ala Ala His Gln Lys Ala Val Val Glu Thr Leu Leu Gln Glu Asp Pro
100 105 110

35 Trp Arg Val Ala Lys Met Val Lys Ser Tyr Leu Gln Gln His Asn Ile
115 120 125

Pro Gln Arg Glu Val Val Asp Thr Thr Gly Leu Asn Gln Ser His Leu
130 135 140

40 Ser Gln His Leu Asn Lys Gly Thr Pro Met Lys Thr Gln Lys Arg Ala
145 150 155 160

Ala Leu Tyr Thr Trp Tyr Val Arg Lys Gln Arg Glu Val Ala Gln Gln
165 170 175

45 Phe Thr His Ala Gly Gln Gly Leu Ile Glu Glu Pro Thr Gly Asp
180 185 190

Glu Leu Pro Thr Lys Lys Gly Arg Arg Asn Arg Phe Lys Trp Gly Pro
195 200 205

50 Ala Ser Gln Gln Ile Leu Phe Gln Ala Tyr Glu Arg Gln Lys Asn Pro
210 215 220

55 Ser Lys Glu Glu Arg Glu Thr Leu Val Glu Glu Cys Asn Arg Ala Glu
225 230 235 240

Cys Ile Gln Arg Gly Val Ser Pro Ser Gln Ala Gln Gly Leu Gly Ser
245 250 255

60 Asn Leu Val Thr Glu Val Arg Val Tyr Asn Trp Phe Ala Asn Arg Arg
260 265 270

Lys Glu Glu Ala Phe Arg His Lys Leu Ala Met Asp Thr Tyr Ser Gly
275 280 285

5 Pro Pro Pro Arg Ala Arg Pro Gly Thr Cys Ala Ala Arg Ser Gln Leu
290 295 300

Pro Trp Pro Ala Ser Thr Cys Pro Leu Pro
305 310

10

(2) INFORMATION FOR SEQ ID NO:7:

15 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3236 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

20

25 (ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: 988
- (D) OTHER INFORMATION: /mod_base= OTHER

/note= "N = A, C, G, or T"

30 (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: join(24..986, 990..1271)

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CGTGGCCCTG TGGCAGCCGA GCC ATG GTT TCT AAA CTG AGC CAG CTG CAG 50
Met Val Ser Lys Leu Ser Gln Leu Gln
1 5

40 ACG GAG CTC CTG GCG GCC CTG CTC GAG TCA GGG CTG AGC AAA GAG GCA 98
Thr Glu Leu Leu Ala Ala Leu Leu Glu Ser Gly Leu Ser Lys Glu Ala
10 15 20 25

45 CTG ATC CAG GCA CTG GGT GAG CCG GGG CCC TAC CTC CTG GCT GGA GAA 146
Leu Ile Gln Ala Leu Gly Glu Pro Gly Pro Tyr Leu Leu Ala Gly Glu
30 35 40

50 GGC CCC CTG GAC AAG GGG GAG TCC TGC GGC GGC GGT CGA GGG GAG CTG 194
Gly Pro Leu Asp Lys Gly Glu Ser Cys Gly Gly Arg Gly Glu Leu
45 50 55

45

50

GCT GAG CTG CCC AAT GGG CTG GGG GAG ACT CGG GGC TCC GAG GAC GAG 242
Ala Glu Leu Pro Asn Gly Leu Gly Glu Thr Arg Gly Ser Glu Asp Glu
60 65 70

55

ACG GAC GAC GAT GGG GAA GAC TTC ACG CCA CCC ATC CTC AAA GAG CTG 290
Thr Asp Asp Asp Gly Glu Asp Phe Thr Pro Pro Ile Leu Lys Glu Leu
75 80 85

60

GAG AAC CTC AGC CCT GAG GAG GCG GCC CAC CAG AAA GCC GTG GTG GAG 338
Glu Asn Leu Ser Pro Glu Glu Ala Ala His Gln Lys Ala Val Val Glu
90 95 100 105

	ACC CTT CTG CAG GAG GAC CCG TGG CGT GTG GCG AAG ATG GTC AAG TCC Thr Leu Leu Gln Glu Asp Pro Trp Arg Val Ala Lys Met Val Lys Ser 110 115 120	386
5	TAC CTG CAG CAG CAC AAC ATC CCA CAG CGG GAG GTG GTC GAT ACC ACT Tyr Leu Gln Gln His Asn Ile Pro Gln Arg Glu Val Val Asp Thr Thr 125 130 135	434
10	GCG CTC AAC CAG TCC CAC CTG TCC CAA CAC CTC AAC AAG GGC ACT CCC Gly Leu Asn Gln Ser His Leu Ser Gln His Leu Asn Lys Gly Thr Pro 140 145 150	482
15	ATG AAG ACG CAG AAG CGG GCC CTG TAC ACC TGG TAC GTC CGC AAG Met Lys Thr Gln Lys Arg Ala Ala Leu Tyr Thr Trp Tyr Val Arg Lys 155 160 165	530
20	CAG CGA GAG GTG GCG CAG CAG TTC ACC CAT GCA GGG CAG GGA GGG CTG Gln Arg Glu Val Ala Gln Gln Phe Thr His Ala Gly Gln Gly Gly Leu 170 175 180 185	578
25	ATT GAA GAG CCC ACA GGT GAT GAG CTA CCA ACC AAG AAG GGG CGG AGG Ile Glu Glu Pro Thr Gly Asp Glu Leu Pro Thr Lys Lys Gly Arg Arg 190 195 200	626
30	AAC CGT TTC AAG TGG GGC CCA GCA TCC CAG CAG ATC CTG TTC CAG GCC Asn Arg Phe Lys Trp Gly Pro Ala Ser Gln Gln Ile Leu Phe Gln Ala 205 210 215	674
35	TAT GAG AGG CAG AAC CCT AGC AAG GAG GAG CGA GAG ACG CTA GTG Tyr Glu Arg Gln Lys Asn Pro Ser Lys Glu Glu Arg Glu Thr Leu Val 220 225 230	722
40	GAG GAG TGC AAT AGG GCG GAA TGC ATC CAG AGA GGG GTG TCC CCA TCA Glu Glu Cys Asn Arg Ala Glu Cys Ile Gln Arg Gly Val Ser Pro Ser 235 240 245	770
45	CAG GCA CAG GGG CTG GGC TCC AAC CTC GTC ACG GAG GTG CGT GTC TAC Gln Ala Gln Gly Leu Gly Ser Asn Leu Val Thr Glu Val Arg Val Tyr 250 255 260 265	818
50	AAC TGG TTT GCC AAC CGG CGC AAA GAA GAA GCC TTC CGG CAC AAG CTG Asn Trp Phe Ala Asn Arg Arg Lys Glu Ala Phe Arg His Lys Leu 270 275 280	866
55	GCC ATG GAC ACG TAC AGC GGG CCC CCC CCA GGG CCA GGC CCG GGA CCT Ala Met Asp Thr Tyr Ser Gly Pro Pro Pro Gly Pro Gly Pro Gly Pro 285 290 295	914
60	GCG CTG CCC GCT CAC AGC TCC CCT GGC CTG CCT CCA CCT GCC CTC TCC Ala Leu Pro Ala His Ser Ser Pro Gly Leu Pro Pro Pro Ala Leu Ser 300 305 310	962
	CCC AGT AAG GTC CAC GGT GTG CGC TNT GGA CAG CCT GCG ACC AGT GAG Pro Ser Lys Val His Gly Val Arg Gly Gln Pro Ala Thr Ser Glu 315 320 325	1010
	ACT GCA GAA GTA CCC TCA AGC AGC GGC GGT CCC TTA GTG ACA GTG TCT Thr Ala Glu Val Pro Ser Ser Ser Gly Gly Pro Leu Val Thr Val Ser 330 335 340	1058
	ACA CCC CTC CAC CAA GTG TCC CCC ACG GGC CTG GAG CCC AGC CAC AGC	1106

	Thr Pro Leu His Gln Val Ser Pro Thr Gly Leu Glu Pro Ser His Ser	
345	350	355
360		
5	CTG CTG AGT ACA GAA GCC AAG CTG GTC TCA GCA GCT GGG GGC CCC CTC Leu Leu Ser Thr Glu Ala Lys Leu Val Ser Ala Ala Gly Gly Pro Leu 365	370
	375	
10	1154	
CCC CGT CAG CAC CCT GAC ACT GCA CAG CTT GGA GCA GAC ATC CCC Pro Arg Gln His Pro Asp Ser Thr Ala Gln Leu Gly Ala Asp Ile Pro 380	385	390
15	1202	
AGG CCT CAA CCA GCA GCC CCA GAA CCT CAT CAT GGC CTC ACT TCC TGG Arg Pro Gln Pro Ala Ala Pro Glu Pro His His Gly Leu Thr Ser Trp 395	400	405
15	1250	
GGT CAT GAC CAT CGG GCC TGG TGAGCCTGCC TCCCTGGTC CTACGTTCAC Gly His Asp His Arg Ala Trp 410	415	
20	1301	
CAACACAGGT GCCTCCACCC TGGTCATCGG CCTGGCCTCC ACGCAGGCAC AGAGTGTGCC		1361
GGTCATCAAC AGCATGGGCA GCAGCCTGAC CACCCTGCAG CCCGTCCAGT TCTCCCAGCC		1421
GCTGCACCCCC TCCTACCAGC AGCCGCTCAT GCCACCTGTG CAGAGCCATG TGACCCAGAG		1481
25		
CCCCTTCATG GCCACCATGG CTCAGCTGCA GAGCCCCAC GCCCTCTACA GCCACAAGCC		1541
CGAGGTGGCC CAGTACACCC ACACGGGCCT GCTCCCGAG ACTATGCTCA TCACCGACAC		1601
30		
CACCAACCTG AGCGCCCTGG CCAGCCTCAC GCCCACCAAG CAGGTCTTCA CCTCAGACAC		1661
TGAGGCCTCC AGTGAGTCCG GGCTTCACAC GCCGGCATCT CAGGCCACCA CCCTCCACGT		1721
35		
CCCCAGCCAG GACCCTGCCG GCATCCAGCA CCTGCAGCCG GCCCACCGGC TCAGCGCCAG		1781
CCCCACAGTG TCCTCCAGCA GCCTGGTGCT GTACCAGAGC TCAGACTCCA GCAATGGCCA		1841
40		
GAGCCACCTG CTGCCATCCA ACCACAGCGT CATCGAGACC TTCATCTCCA CCCAGATGGC		1901
CTCTTCCTCC CAGTAACCAC GGCACCTGGG CCCTGGGGCC TGTACTGCCT GCTTGGGGGG		1961
TGATGAGGGC AGCAGCCAGC CCTGCCTGGA GGACCTGAGC CTGCCGAGCA ACCGTGGCCC		2021
45		
TTCCCTGGACA GCTGTGCCTC GCTCCCCACT CTGCTCTGAT GCATCAGAAA GGGAGGGCTC		2081
TGAGGCGCCC CAACCCGTGG AGGCTGCTCG GGGTGACAG GAGGGGGTCG TGGAGAGCTA		2141
GGAGCAAAGC CTGTTCATGG CAGATGTAGG AGGGACTGTC GCTGCTTCGT GGGATACAGT		2201
50		
CTTCTTACTT GGAACTGAAG GGGCGGGCCT ATGACTTGGG CACCCCCAGC CTGGGCCTAT		2261
GGAGAGCCCT GGGACCGCTA CACCACTCTG GCAGCCACAC TTCTCAGGAC ACAGGCCTGT		2321
GTAGCTGTGA CCTGCTGAGC TCTGAGAGGC CCTGGATCAG CGTGGCCTTG TTCTGTCACC		2381
55		
AATGTACCCA CCGGGCCACT CCTCCTGCC CCAACTCCTT CCAGCTAGTG ACCCACATGC		2441
CATTGTACT GACCCCATCA CCTACTCACA CAGGCATTTC CTGGGTGGCT ACTCTGTGCC		2501
60		
AGAGCCTGGG GCTCTAACTG CCTGAGCCCA GGGAGGCCGA AGCTAACAGG GAAGGCAGGC		2561

	AGGGCTCTCC TGGTCTTCCC ATCCCCAGCG ATTCCCTCTC CCAGGCCCA TGACCTCCAG	2621
	CTTTCCCTGTA TTTCTTCCC AGAGCATGAT GCCTCTGAGG CCAGCCTGGC CTCCTGCCTC	2681
5	TACTGGGAAG GCTACTTCGG GGCTGGGAAG TCGTCCTTAC TCCTGTGGGA GCCTCGAAC	2741
	CCGTGCCAAG TCCAGGTCTT GGTGGGGCAG CTCCTCTGTC TCGAGCGCCC TGCAGACCCT	2801
10	GCCCTTGTGTT GGGGCAGGAG TAGCTGAGCT CACAAGGCAG CAAGGCCGA GCAGCTGAGC	2861
	AGGGCCGGGG AACTGGCCAA GCTGAGGTGC CCAGGAGAAG AAAGAGGTGA CCCCAGGGCA	2921
	CAGGAGCTAC CTGTGTGGAC AGGACTAACCA CTCAGAACCC TGGGTGCCTG GCTGGCTGAG	2981
15	GGCAGTTCGC AGCCACCCCTG AGGAGTCTGA GGTCCTGAGC ACTGCCAGGA GGGACAAAGG	3041
	AGCCTGTGAA CCCAGGACAA GCATGGTCCC ACATCCCTGG GCCTGCTGCT GAGAACCTGG	3101
20	CCTTCAGTGT ACCGCGTCTA CCCTGGGATT CAGGAAAAGG CCTGGGGTGA CCCGGCACCC	3161
	CCTGCAGCTT GTAGCCAGCC GGGGCGAGTG GCACGTTTAT TAACTTTTA GTAAAGTCAA	3221
	GGAGAAATGC GGTGG	3236

25 (2) INFORMATION FOR SEQ ID NO:8:

30 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 415 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

35 (ii) MOLECULE TYPE: protein

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met Val Ser Lys Leu Ser Gln Leu Gln Thr Glu Leu Leu Ala Ala Leu
 1 5 10 15

Leu Glu Ser Gly Leu Ser Lys Glu Ala Leu Ile Gln Ala Leu Gly Glu
 20 25 30

Pro Gly Pro Tyr Leu Leu Ala Gly Glu Gly Pro Leu Asp Lys Gly Glu
 35 40 45

45 Ser Cys Gly Gly Arg Gly Glu Leu Ala Glu Leu Pro Asn Gly Leu
 50 55 60

50 Gly Glu Thr Arg Gly Ser Glu Asp Glu Thr Asp Asp Asp Gly Glu Asp
 65 70 75 80

Phe Thr Pro Pro Ile Leu Lys Glu Leu Glu Asn Leu Ser Pro Glu Glu
 85 90 95

55 Ala Ala His Gln Lys Ala Val Val Glu Thr Leu Leu Gln Glu Asp Pro
 100 105 110

Trp Arg Val Ala Lys Met Val Lys Ser Tyr Leu Gln Gln His Asn Ile
 115 120 125

60 Pro Gln Arg Glu Val Val Asp Thr Thr Gly Leu Asn Gln Ser His Leu

	130	135	140
	Ser Gln His Leu Asn Lys Gly Thr Pro Met Lys Thr Gln Lys Arg Ala		
5	145	150	155
	Ala Leu Tyr Thr Trp Tyr Val Arg Lys Gln Arg Glu Val Ala Gln Gln		
	165	170	175
10	Phe Thr His Ala Gly Gln Gly Gly Leu Ile Glu Glu Pro Thr Gly Asp		
	180	185	190
	Glu Leu Pro Thr Lys Lys Gly Arg Arg Asn Arg Phe Lys Trp Gly Pro		
	195	200	205
15	Ala Ser Gln Gln Ile Leu Phe Gln Ala Tyr Glu Arg Gln Lys Asn Pro		
	210	215	220
20	Ser Lys Glu Glu Arg Glu Thr Leu Val Glu Glu Cys Asn Arg Ala Glu		
	225	230	235
	Cys Ile Gln Arg Gly Val Ser Pro Ser Gln Ala Gln Gly Leu Gly Ser		
25	245	250	255
	Asn Leu Val Thr Glu Val Arg Val Tyr Asn Trp Phe Ala Asn Arg Arg		
	260	265	270
30	Lys Glu Glu Ala Phe Arg His Lys Leu Ala Met Asp Thr Tyr Ser Gly		
	275	280	285
35	Pro Pro Pro Gly Pro Gly Pro Gly Pro Ala Leu Pro Ala His Ser Ser		
	290	295	300
40	Pro Gly Leu Pro Pro Pro Ala Leu Ser Pro Ser Lys Val His Gly Val		
	305	310	315
	Arg Gly Gln Pro Ala Thr Ser Glu Thr Ala Glu Val Pro Ser Ser Ser		
45	325	330	335
	Gly Gly Pro Leu Val Thr Val Ser Thr Pro Leu His Gln Val Ser Pro		
	340	345	350
	Thr Gly Leu Glu Pro Ser His Ser Leu Leu Ser Thr Glu Ala Lys Leu		
	355	360	365
50	Val Ser Ala Ala Gly Gly Pro Leu Pro Arg Gln His Pro Asp Ser Thr		
	370	375	380
	Ala Gln Leu Gly Ala Asp Ile Pro Arg Pro Gln Pro Ala Ala Pro Glu		
	385	390	395
	Pro His His Gly Leu Thr Ser Trp Gly His Asp His Arg Ala Trp		
	405	410	415

(2) INFORMATION FOR SEQ ID NO:9:

- 5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

- 10 (ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 7
(D) OTHER INFORMATION: /mod_base= OTHER

15 /note= "N = A, C, G, or T"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

20 GTTAATNATT ACC

13

(2) INFORMATION FOR SEQ ID NO:10:

- 25 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

TACACCACTC TGGCAGCCAC ACT

23

35 (2) INFORMATION FOR SEQ ID NO:11:

- 40 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

CGGTGGGTAC ATTGGTGACA GAAC

24

50 (2) INFORMATION FOR SEQ ID NO:12:

- 55 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

60 GGCAGGCAAA CGCAACCCAC G

21

(2) INFORMATION FOR SEQ ID NO:13:

- 5 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 21 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GAAGGGGGGC TCGTTAGGAG C

21

15 (2) INFORMATION FOR SEQ ID NO:14:

- 20 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 21 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

CATGCACAGT CCCCCACCCCTC A

21

30 (2) INFORMATION FOR SEQ ID NO:15:

- 35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 21 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

CTTCCAGCCC CCACCTATGA G

21

45 (2) INFORMATION FOR SEQ ID NO:16:

- 50 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 21 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

GGGCAAGGTC AGGGGAATGG A

21

60 (2) INFORMATION FOR SEQ ID NO:17:

- 55 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 21 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

CAGCCCAGAC CAAACCAGCA C

21

5

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

15

CAGAACCCCTC CCCTTCATGC C

21

20

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

25

GGTGACTGCT GTCAATGGGA C

21

30

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

40

GGCAGACAGG CAGATGGCCT A

21

45

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

55

GCCTCCCTAG GGACTGCTCC A

21

(2) INFORMATION FOR SEQ ID NO:22:

- 5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

TGGAGCAGTC CCTAGGGAGG C

21

15 (2) INFORMATION FOR SEQ ID NO:23:

- 20 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GTTGCCCAT GAGCCTCCCA C

21

30 (2) INFORMATION FOR SEQ ID NO:24:

- 35 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GGTCTGGGC AGGGGTGGGA T

21

45 (2) INFORMATION FOR SEQ ID NO:25:

- 50 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

CTGCAATGCC TGCCAGGCAC C

21

60 (2) INFORMATION FOR SEQ ID NO:26:

- 55 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

CCCCTGCATC CATTGACAGC C

21

5

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

20 GAGGCCTGGG ACTAGGGCTG T

21

25 (2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

- 30 (A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

35 CTCTGTCACA GGCGGAGGGA G

21

40 (2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

- 45 (A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

55 CCTGTGACAG AGCCCCTCAC C

21

45 (2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

- 50 (A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

CGGACAGCAA CAGAAGGGT G

21

(2) INFORMATION FOR SEQ ID NO:31:

- 5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

CAGAGCCCCCT CACCCCCACA T

21

15 (2) INFORMATION FOR SEQ ID NO:32:

- 20 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

GTACCCCTAG GGACAGGCAG G

21

30 (2) INFORMATION FOR SEQ ID NO:33:

- 35 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

ACCCCCCAAG CAGGCAGTAC A

21

45 (2) INFORMATION FOR SEQ ID NO:34:

- 50 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 671 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

- 55 (ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 104..217

60 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

GCAGAGAGGG CACTGGGAGG AGGCAGTGGG AGGGCGGAGG GCGGGGGCCT TCGGGGTGGG

60

CGCCCCAGGGT AGGGCAGGTG GCCGGCGCGT GGAGGCAGGG AGA ATG CGA CTC TCC

115

Met Arg Leu Ser
1

5 AAA ACC CTC GTC GAC ATG GAC ATG GCC GAC TAC AGT GCT GCA CTG GAC 163
Lys Thr Leu Val Asp Met Asp Met Ala Asp Tyr Ser Ala Ala Leu Asp
5 10 15 20

10 CCA GCC TAC ACC ACC CTG GAA TTT GAG AAT GTG CAG GTG TTG ACG ATG 211
Pro Ala Tyr Thr Thr Leu Glu Phe Glu Asn Val Gln Val Leu Thr Met
25 30 35

15 GGC AAT GGTAGGTGGG GGCAGATGTG CCCAGGTGTG CCAGTGCCCC CAGGTGTGCC 267
Gly Asn

20 TGGGTCCAGG AGCAGATCTT TGGCACTCAA CTTTGGGGTG GGAGGAGAAT GATACAAAAT 327

25 GGTAGGTTGG TCCTACAGGC CAGCACAGGT GTTGCCAAGT GAAGCCCAGT TGCCCAGGCA 387

30 CAGTGATCAC AGGCATTCTG GGTGAAGGGA GGCCTGCAAG GGCCAATTTC CAGCAAAAGT 447

35 CGATCCCGGC TATTCCCTCCC AGGCCCTTCC AGTCCTCACT GCCTCACAGT GGCTCTGCTT 507

40 GGCCTTGGC ACAGTGACAT GATGGTGAGC TCCCCCTTGG TGCCCAGCTC CAGCGATTCA 567

45 GCCCAGCACG GCCCCTTCGT GAACCCCTTG GGCCTAGGTT CAGAGAGACG GCAAGGGATG 627

50 TTGTATCCCT GGAGATGGTG GTTGGAGACA TAACCGCATT TCTC 671

(2) INFORMATION FOR SEQ ID NO:35:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 38 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

Met Arg Leu Ser Lys Thr Leu Val Asp Met Asp Met Ala Asp Tyr Ser
1 5 10 15

45 Ala Ala Leu Asp Pro Ala Tyr Thr Thr Leu Glu Phe Glu Asn Val Gln
20 25 30

Val Leu Thr Met Gly Asn
35

(2) INFORMATION FOR SEQ ID NO:36:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 796 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

60

5 (ix) FEATURE:

- (A) NAME/KEY: CDS
(B) LOCATION: join(286..312, 316..375)

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

TGGATGTTTG TACATGTGTG CTGTGTGTGC GGGTCATAGA GCACATGTGT TTGTGCATGC 60
10 GGACCTGTTG GAGTGCCCTG TTCTTCCTGC ATCTTATCC TGTATGGCGC TTTTGTCTGT 120
TGCCTCATATT TGTACCTGCT GTGTATATAT GCAGTCCCT GTGCTGCAGG CGGGGGTCAG 180
15 CGGTCTCTGG TGTGCACGAC TGCACAGACC CAAATGCAGG ACTCTGTTGT TGCCACTCAC 240
CAAGTGAGAT TCATATCAGC AACATGTCCG TTTGTCTCTG AGCAG ATT TGT TGC 294
Ile Cys Cys
1
20 CGC TGC GTC TCG CCA GAT TGA GGC ATC CCC TCC GAC ATC ACT GGA GCA 342
Arg Cys Val Ser Pro Asp Gly Ile Pro Ser Asp Ile Thr Gly Ala
5 10 15
25 TAT CTG GAG GGG TGG ACA GTT CTC CAC AGG GAG GTAGGGAAA AGAGGAGGCC 395
Tyr Leu Glu Gly Trp Thr Val Leu His Arg Glu
20 25
30 CGGAAACCCC TCCTGGAGGG AAGAGCCCCA TCGGTCCCAG GCCAGCCTCA GAGGAGAGGG 455
35 GGCAGGCAGC TGGCTGAGGT CAGCCTGCCA CCCTGCTTCC TTCTGTGTCT TGGAGCCACT 515
CAGCCAGTAT GAGGCTGCAG CTCCAGCTGA GGTCTGGAAT CTTGTGGTCA GCTCAGCTAG 575
40 GGTGAGGAGG CAGCTGCTGG GCACTGCTTG TTGTCAGCTC AGCAGGTGCT CACCTGCC 635
TGCCGTCCAG TCACGTGTGA CCTTGGGCAT GTCACCTCCC CTATCCTGGC TTCTGTATCT 695
TCTACAAAAC AGGCTTCATT CCCCCAGGCC TGCTGGCTGG ACGGCTTTA GGCCTGTCTG 755
AGGACCACGC CAGGAGCGCA AGGCAAAAC ACACCAGAGA T 796

45 (2) INFORMATION FOR SEQ ID NO:37:

50 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

55 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

Ile Cys Cys Arg Cys Val Ser Pro Asp Gly Ile Pro Ser Asp Ile Thr
1 5 10 15
Gly Ala Tyr Leu Glu Gly Trp Thr Val Leu His Arg Glu
20 25

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 634 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 326..499

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

(2) INFORMATION FOR SEO ID NO: 39:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 58 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

5 Thr Arg Pro His Gln Lys Ala Pro Thr Ser Thr Arg Pro Thr Ala Trp
1 5 10 15

Val Ser Ala Pro Cys Val Pro Ser Ala Gly Thr Gly Pro Arg Ala Asn
20 25 30

10 Thr Thr Val Pro Arg Ala Val Thr Ala Ala Arg Ala Ser Ser Gly Gly
35 40 45

Ala Cys Gly Arg Thr Thr Cys Thr Pro Ala
50 55

15 (2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:

- 20 (A) LENGTH: 458 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

25 (ix) FEATURE:

- (A) NAME/KEY: CDS
(B) LOCATION: join(171..173, 177..265)

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

AGAGAGTTCA TAGCACCTT CCAGCTCCTG GTGGGTTCAA GAGAGAACTC CCGGGATGAA 60

GAGATGAGAG CACTGAGGTT GGGGGGTCAA CTGGATAGCC AGGGCCCTAG TTCTGTCCTA 120

AGAGGAGGAA GTTGTGTCTT CTCCATCCAA CCATCCAAAG CCCTCCCCAG ATT 173

 Ile
 1

40 TAG CCG GCA GTG CGT GGT GGA CAA AGA CAA GAG GAA CCA GTG CCG CTA 221

 Pro Ala Val Arg Gly Gly Gln Arg Gln Glu Glu Pro Val Pro Leu
 5 10 15

45 CTG CAG GCT CAA GAA ATG CTT CCG GGC TGG CAT GAA GAA GGA 263

Leu Gln Ala Gln Glu Met Leu Pro Gly Trp His Glu Glu Gly
 20 25 30

AGGTGAGCCT CGGCCCTCCC CGCCCCACCA CCACTGCCCT ACCTGCACCC ACAGCTCCCC 323

50 GACAGTCATT TACAACGTGA GCCACACTTT ATGACTCAGT GGCAGGCCCC AGGGTGACTG 383

GCTAATGGCT GAGAAGAGGG AGGGCCTGGA AATCTGACCA TAGGGAGCGG CTGGGCTTGG 443

TCTTGAGAAA GATTC 458

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

Ile Pro Ala Val Arg Gly Gly Gln Arg Gln Glu Glu Pro Val Pro Leu
1 5 10 15

Leu Gln Ala Gln Glu Met Leu Pro Gly Trp His Glu Glu Gly
20 25 30

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 662 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
(B) LOCATION: 84..188

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

TCCCACTCCT CATCAGTCAC AGACACCCCC ACCCCCTACT CCATCCCTGT TCTCCCTCCT 60

CACCTCTCTG TGCCTCCTCA CAG CCG TCC AGA ATG AGC GGG ACC GGA TCA
Pro Ser Arg Met Ser Gly Thr Gly Ser
1 5 110

GCA CTC GAA GGT CAA GCT ATG AGG ACA GCA GCC TGC CCT CCA TCA ATG
Ala Leu Glu Gly Gln Ala Met Arg Thr Ala Ala Cys Pro Pro Ser Met
10 15 20 25 158

CGC TCC TGC AGG CGG AGG TCC TGT CCC GAC AGGTACCGGG GTGATCCTGC 208
Arg Ser Cys Arg Arg Ser Cys Pro Asp
30 35

CACCCACCCA GGGGATCCCC CACACTACAG AGGAGCTCAC CTCCCTCCACC TCCATTCTCC 268

CCAGGCCAGGC CCTGGAGCAG CTGACGGGAG GGGCCTCAGA TATTACAGAA GGGACACTGA 328

GTGCGGTTTC ACATGGCCCA GTTTGCAGCA AGGGCAGGAA TCGAACCTGG CGCCCTGGGG 388

55 CACTTTCTAA TTCATCCTAC TGCCTGCATC CCACAGGCCA AGCAGAGTCT TCACCTTCAC 448

TGAGGGCCTG CGATCAGCTC AGCTCCGAGA GAACAGAGCA GTGGCTCAGT GGAGAGAGGT 508

GGCAAAGTGG GGCCCAGCCC TTCCCTTGCT GAGTGACCTT GGGCAAGTCA CAGCACCTCT 568

60 CTGAGGCCATG GTTGCCTCAT TGTCAAGAAAA GGATGATGAT TTTTGCCCT GCTTCTCCTC 628

TAAGGCTGAC AGACTCCTTG GGGCTCTAAA GCTG

662

5 (2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

15 Pro Ser Arg Met Ser Gly Thr Gly Ser Ala Leu Glu Gly Gln Ala Met
1 5 10 15

20 Arg Thr Ala Ala Cys Pro Pro Ser Met Arg Ser Cys Arg Arg Arg Ser
20 25 30

25 Cys Pro Asp
35

30 (2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 647 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

35 (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 185..340

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

45 TTCTCCCTCA TCCCTGCCTC CTCCCTCCCT CCGTTTTAC CCTGAGCTTC CTTCAGAGCT 60

AGAGGGCACC CACTATCCAG CCCCTCCCC ACATCTGATT CCAGGGAGGG GGCTCTGTGC 120

50 AGGGGACAGA GAATGCGGGA GGGCCCGGAC ATCTCCAGCA TTTTCTTCCC TGTATCTCTC 180

GAAG ATC ACC TCC CCC GTC TCC GGG ATC AAC GGC GAC ATT CGG GCG AAG
Ile Thr Ser Pro Val Ser Gly Ile Asn Gly Asp Ile Arg Ala Lys
1 5 10 15 229

55 AAG ATT GCC AGC ATC GCA GAT GTG TGT GAG TCC ATG AAG GAG CAG CTG 277
Lys Ile Ala Ser Ile Ala Asp Val Cys Glu Ser Met Lys Glu Gln Leu
20 25 30

60 CTG GTT CTC GTT GAG TGG GCC AAG TAC ATC CCA GCT TTC TGC GAG CTC 325
Leu Val Leu Val Glu Trp Ala Lys Tyr Ile Pro Ala Phe Cys Glu Leu
35 40 45

CCC CTG GAC GAC CAG GTGAGGATGG GCGTGGATGG TGGGCAGTAG TGGGCAGTGG 380
Pro Leu Asp Asp Gln

50

GC	GGGGCAGC	CAGGGGGCTG	CTGGCCCACC	TGGGATATAG	CCGTGGACTG	GCTTGATT	440			
5	TTT	TTTA	ACAAAATATG	TAGTGCACAC	ACGTGTCTGA	AACTTTAAAT	CACCTTACAA	500		
			ATATTAAC	TC	CAACAACT	CTATGAGGTA	GGTACTAAGG	TACTATT	560	
10			ACTGCCAT	CT	CATAGGTGAG	AGATTGGGC	ACAGAGAGGT	TAAGTAACCT	GCTCAAGGTC	620
			ACATAGCTAC	TATCCAGCAT	AGCTGGG				647	

(2) INFORMATION FOR SEQ ID NO:45:

- 15 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 52 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

20 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

Ile	Thr	Ser	Pro	Val	Ser	Gly	Ile	Asn	Gly	Asp	Ile	Arg	Ala	Lys	Lys
1					5				10					15	
Ile	Ala	Ser	Ile	Ala	Asp	Val	Cys	Glu	Ser	Met	Lys	Glu	Gln	Leu	Leu
					20				25					30	
Val	Leu	Val	Glu	Trp	Ala	Lys	Tyr	Ile	Pro	Ala	Phe	Cys	Glu	Leu	Pro
					35			40				45			
Leu	Asp	Asp	Gln												
			50												

(2) INFORMATION FOR SEQ ID NO:46:

- 40 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 844 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

- 45 (ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 429..515

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

AT	TTT	TACAA	AGCACCC	TT	CT	ATAATT	CTCC	ATAGCT	GGTC	CATGG	GT	GGG	AAT	TT	GGG	AC	60	
55																		120
			CCACAG	TT	TT	TT	GGAA	CT	TT	GG	AT	CT	CAA	AA	AG	AA		
			AAG	CAC	A	CAG	AAT	GTT	GCT	AC	AGT	TT	CAT	CAG	CAC	GA	AG	180
60																		240
			AG	TT	CT	TG	CC	AAG	GAC	AC	AG	TT	CAA	GG	AC	AG	GT	

TGAGCACATG TTCTTCCCC TTCCAGGTT CTAGTTTAT GGGTACTAGT TTTATGATGC 300
 CCATTCACA GTTCAGGCAG GTAGAGGCAG AGGGGAGCAT TAAGCTGACT TGCCCAGCGT 360
 5 CACTGAGTTG GCTACGGGCA GCCTTCCCAA GGGTACAGAT GGCAAACACT GTTCCTTATC 420
 TCTTCAG GTG GCC CTC AGA GCC CAT GCT GGC GAG CAC CTG CTG CTC 470
 Val Ala Leu Leu Arg Ala His Ala Gly Glu His Leu Leu Leu
 1 5 10

10 GGA GCC ACC AAG AGA TCC ATG GTG TTC AAG GAC GTG CTG CTC CTA 515
 Gly Ala Thr Lys Arg Ser Met Val Phe Lys Asp Val Leu Leu Leu
 15 20 25

15 GGTGAGGCAG CTGCCTGCC TGGCCAGGGC TCCAGGGAGG GTATGCCTAG CATGGCACTC 575
 ACCCAGGCAA GGAGATTACAT ATGGTGGCAT GCAAGGGTGA GGGAGACTAG TCAGGAGTGG 635
 20 CCCTGTCCTC AGGCTTGAT TGGAGGGCTC CAGGACTCAG TTTTCAACTG GGTACCCAC 695
 TCAGATGCAA GGAAATGTGG ATGCAAGTCA CCAAATTCCC AGCATTGAAG TCAGAGCACG 755
 25 ATCAGGGTTA TCCCTGGAAT TACCTGTGCA TCCTTTTTC TTTTGACAGA GTCTTGCTCT 815
 GTCACTCAGG CTGGAGTGCA ATGATGTGA 844

30 (2) INFORMATION FOR SEQ ID NO:47:

- 35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 29 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

40 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

45 Val Ala Leu Leu Arg Ala His Ala Gly Glu His Leu Leu Leu Gly Ala
 1 5 10 15

Thr Lys Arg Ser Met Val Phe Lys Asp Val Leu Leu Leu
 20 25

50 (2) INFORMATION FOR SEQ ID NO:48:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 937 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

55 (ix) FEATURE:

- (A) NAME/KEY: CDS
 (B) LOCATION: join(485..529, 533..640)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

GCAACACTAG TATTTAATA TAACAATGCT ATGAGGGAGC TCGATTATT ATCCTCATCT 60
5 TATAGATAAG AAAACTGAGG CACAGAGAGG TTAAGTAACT TATCCAAC TAACCAGCTA 120
TCAGGGGCAG AGCCATTAA GCAGGGCAGT GCAGTCCAG AATCTGGTCC TTTAACCTTG 180
10 ATGCTTTGGT GCCTATCAGG TGACCTTGGA ATGTCATCGA TCTTGTGAGT CATGTTGGTA 240
AATGGAGCTT GGGTCATGTG AAAGAGGTCC TAGAAAGCCA AGTTCCAAGC TCAGCCGGAT 300
GACTCAAGGC AGCTTATCTT CTGAATCTGG GCCTCAGCTT CCTTACCTGT GAAATGGGAG 360
15 TCACCATCCC TGCAGGTCCCT CCTCCCACAG GCACCAGCTA TCTTGCCAAC TTAAAAGCCA 420
AAACTAGAGG AGAGGGGTCA ACCCAAAGTG ACTTCCCAC TCTCCCTCCCT CCCAACCCCTT 480
20 CCAG GCA ATG ACT ACA TTG TCC CTC GGC ACT GCC CGG AGC TGG CGG AGA 529
Ala Met Thr Thr Leu Ser Leu Gly Thr Ala Arg Ser Trp Arg Arg
1 5 10 15
TGA GCC GGG TGT CCA TAC GCA TCC TTG ACG AGC TGG TGC TGC CCT TCC 577
25 Ala Gly Cys Pro Tyr Ala Ser Leu Thr Ser Trp Cys Cys Pro Ser
20 25 30
AGG AGC TGC AGA TCG ATG ACA ATG AGT ATG CCT ACC TCA AAG CCA TCA 625
Arg Ser Cys Arg Ser Met Thr Met Ser Met Pro Thr Ser Lys Pro Ser
35 40 45
TCT TCT TTG ACC CAG GTACAGTGCA CACCTCCTAA GCCATCCCTG ACTCTCTCTC 680
Ser Ser Leu Thr Gln
50
CAGAACGCTC TGCCAGACTT CTCCTATTGG GTTCTGTACA CTGAGTTCAC AGCCTCATCT 740
35
CATGTTAACG ACAGCCAGGA GAGGCCGTT TCATTTAAC AATGAGGCAA GTCAAGATTT 800
40 GAAGAGACAA TATGGCCGGG CGCAGTGGCT CACACCTGTA ATCCCACATCAC TTTGGGAGGC 860
TGAGGCAGGGC GGATCACCTG AGGTCAGGGG TCAAGATGAG CCTGGCTAAC ATGGAGAAC 920
CCCCATCTCTA CTTAAAAA
937

45

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:

- 50 (A) LENGTH: 51 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

60

Ala Met Thr Thr Leu Ser Leu Gly Thr Ala Arg Ser Trp Arg Arg Ala
1 5 10 15
Gly Cys Pro Tyr Ala Ser Leu Thr Ser Trp Cys Cys Pro Ser Arg Ser
20 25 30

Cys Arg Ser Met Thr Met Ser Met Pro Thr Ser Lys Pro Ser Ser Ser
35 40 45

5 Leu Thr Gln
50

10 (2) INFORMATION FOR SEQ ID NO:50:

- 15 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 978 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

20 (ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: join(376..387, 391..432, 436..534, 538..610)

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

GTGGCTCTGC CAACAACTGG CTGTGCGACC CAGGACAAGT CCTATCTTG CACTGTGTCT 60
GGGTTTCCCC GTGTGTAAGA TGAGGCGGTT GCTAGGTGCT TATTGGATGC ATTCCCTCAAG 120
TCCCGCCCTC CATCTCCTAT TCCCCTCTCT TCTGGTTTAG TGCTTTAGGA AATGTGGCAG 180
AAATCTTTTT CTGCCTGTGT CTAGGAAATC ATAATTCATG CTGGCGTACC CTGGTTGTTG 240
AGGTCCCTGA ATCCTGTGC CCACACTGCT GAAGACTCCT TGTGTGACAC AAGTCAGGGG 300
ACATCTGGGT CTTGACTCCC CAGATGCTCC AGGTGGACCC TGCTGCCCTC CCTTGCCCCAC 360
CCTCTTCCAT TGTAG ATG CCA AGG GGC TGA GCG ATC CAG GGA AGA TCA AGC 411
Met Pro Arg Gly Ala Ile Gln Gly Arg Ser Ser
1 5 10
40 GGC TGC GTT CCC AGG TGC AGG TGA GCT TGG AGG ACT ACA TCA ACG ACC 459
Gly Cys Val Pro Arg Cys Arg Ala Trp Arg Thr Thr Ser Thr Thr
15 20 25
45 GCC AGT ATG ACT CGC GTG GCC GCT TTG GAG AGC TGC TGC TGC TGC 507
Ala Ser Met Thr Arg Val Ala Ala Leu Glu Ser Cys Cys Cys Cys Cys
30 35 40
50 CCA CCT TGC AGA GCA TCA CGT GGC AGA TGA TCG AGC AGA TCC AGT TCA 555
Pro Pro Cys Arg Ala Ser Arg Gly Arg Ser Ser Arg Ser Ser Ser
45 50 55
55 TCA AGC TCT TCG GCA TGG CCA AGA TTG ACA ACC TGT TGG AGG AGA TGC 603
Ser Ser Ser Ala Trp Pro Arg Leu Thr Thr Cys Trp Arg Arg Cys
60 65 70
60 TGC TGG GAGGTCCGTG CCAAGCCCAG GAGGGGCGGG GTTGGATTGG GGACTCCCCA 659
Cys Trp
75
GGAGACAGGC CTCACACAGT GAGCTCACCC CTCAGCTCCT TGGCTTCCCC ACTGTGCCGC 719

TTTGGGCAAG TTGCTTAACC TGTCTGTGCC TCAGTTCCCT CACCAGAAAA ATGGGAACAA	779
5 GGCAATGGTC TATTGTTCA GGCACCGAGA ACCTAGCACG TGCCAGTCAC TGTTCTAAGT	839
GCTGGCAATT CAGCAAAGAA CAAGATCTT GCCCTCGGGG AGGCTGTGTG TGTGTGATAT	899
10 GTATGGATGC GTGGATATCT GTGTATATGC CCGTATGTGC GTGCATGTGT ATATAAAGCC	959
TCACATTTA TGATTTGA	978

20 (2) INFORMATION FOR SEQ ID NO:51:

- 25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 75 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

30 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

Met Pro Arg Gly Ala Ile Gln Gly Arg Ser Ser Gly Cys Val Pro Arg	
1 5 10 15	
Cys Arg Ala Trp Arg Thr Thr Ser Thr Thr Ala Ser Met Thr Arg Val	
20 25 30	
Ala Ala Leu Glu Ser Cys Cys Cys Cys Pro Pro Cys Arg Ala Ser	
35 40 45	
Arg Gly Arg Ser Ser Arg Ser Ser Ser Ser Ser Ala Trp Pro	
50 55 60	
Arg Leu Thr Thr Cys Trp Arg Arg Cys Cys Trp	
65 70 75	

40 (2) INFORMATION FOR SEQ ID NO:52:

- 45 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 984 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

50 (ix) FEATURE:

- (A) NAME/KEY: CDS
 (B) LOCATION: join(443..490, 494..595)

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

GGGACACATA GATGCTATAA GTAGGTCAGT TGGCTGCAGC AGAGATGTGG GGGATGAGGC	60
TGAAAGGTGA GGCAGGACCA AATGGTTGAA GGACTTGAC TCCAAGGAGC TTTGAGAGCC	120
60 ATTGATTACA TCCATTATGT TACTATGTGA CCAATACATT ACTCATTAGA ACATTTACGT	180

	GATCTCAGAG CTTCCCTTATA TGCACCTTGT TCCTTCAAC TCACCTTGT TCTCTGGTT	240
	TTTTGGGGTC CTCTAACAC CCTCATGAAG TCTATAGATG GGAATGGTAC ACCCTAGTT	300
5	ACTAACCCAG GAATAGGTAC CCAACAGGCA CTGCCAATAT TGGATGGCT GGTTGATTGG	360
	CCACGCCTGA GGAAGATGGC GTCCCAAGGC CTGAGGTCTG CATCCCAGAC TCTCCATCCT	420
10	GATCGACCTT CTCTACCTGC AG GGT CCC CCA GCG ATG CAC CCC ATG CCC ACC Gly Pro Pro Ala Met His Pro Met Pro Thr 1 5 10	472
15	ACC CCC TGC ACC CTC ACC TGA TGC AGG AAC ATA TGG GAA CCA ACG TCA Thr Pro Cys Thr Leu Thr Cys Arg Asn Ile Trp Glu Pro Thr Ser 15 20 25	520
20	TCG TTG CCA ACA CAA TGC CCA CTC ACC TCA GCA ACG GAC AGA TGT GTG Ser Leu Pro Thr Gln Cys Pro Leu Thr Ser Ala Thr Asp Arg Cys Val 30 35 40	568
25	AGT GGC CCC GAC CCA GGG GAC AGG CAG GTGGCAAAC TCTGGATTT Ser Gly Pro Asp Pro Gly Asp Arg Gln 45 50	615
30	TACCTTGCAA AGGGTGAGGA TGGGGCTTAA GACAGGAGGC AGGAGAAAGT GGAGTCTAGA	675
35	AGGTAGAACCC AGGATGCAAC AGTTTCTGG GTTCCAGGGT AGGGAATAAA GGGCAAGATT	735
40	GTCCATTGTT TGAGGCTGTT TATTCAGTAA GGTGACTGAC AGCCTTACT GAATGAAGCC	795
45	ATTGTTGGGA TGAGGCAATC CACTGGATGA GGTAACCCAT TGGGTGAAGA TGTCTGGGT	855
50	GAGAATTCCA TTAGTTGACA TTGTCCATTA AGTAAAAGTG GTCATTGAAG TAAGGCTGCA	915
55	CAGTTGGGTA AGGCTATCCA TTAGACATTA GATGAGACTA CCCATTGGGT CAGGATGTCT	975
60	GCTGGGCTA	984

40 (2) INFORMATION FOR SEQ ID NO:53:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 50 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

45 (ii) MOLECULE TYPE: protein

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

Gly	Pro	Pro	Ala	Met	His	Pro	Met	Pro	Thr	Thr	Pro	Cys	Thr	Leu	Thr
1														15	
Cys	Arg	Asn	Ile	Trp	Glu	Pro	Thr	Ser	Ser	Leu	Pro	Thr	Gln	Cys	Pro
Leu	Thr	Ser	Ala	Thr	Asp	Arg	Cys	Val	Ser	Gly	Pro	Asp	Pro	Gly	Asp
35	40													45	
Arg	Gln														
	50														

(2) INFORMATION FOR SEQ ID NO:54:

5 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1103 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

10 (ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: join(289..429, 433..477, 481..492, 496..603, 607
 ..630, 634..750, 754..810, 814..843, 847..1023,
 1027..1071, 1075..1103)

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

25	TTGGGAGAA GCAGTCCAAG TCTGCATATC AAATAATGA TGGAGGAGAT GGGTGGTAGG	60
30	ACCTTCCAGA CCTCATAAAAA CTTAGGCTTT ATGATCTGGG ACTCACAGAA GGTTGAGCAA	120
35	TAAAAGACCT TAGGGATTAT CTGGCTTAAT TAATTCTCTC ATTTTATAGA GGAAGAAATT	180
40	AAGTCAAGGT GGGGCAGGGT GGGAGGGGAG AACTTCCCCG GGGCTCTTCA TTTACTCCCA	240
45	CAAAGGCTGG AATTTGAGC AGCCCCTGTC TGTCTGTTG TCCTTCCA GCC ACC CCT Ala Thr Pro 1	297
50	GAG ACC CCA CAG CCC TCA CCG CCA GGT GGC TCA GGG TCT GAG CCC TAT Glu Thr Pro Gln Pro Ser Pro Pro Gly Gly Ser Gly Ser Glu Pro Tyr 5 10 15	345
55	AAG CTC CTG CCG GGA GCC GTC GCC ACA ATC GTC AAG CCC CTC TCT GCC Lys Leu Leu Pro Gly Ala Val Ala Thr Ile Val Lys Pro Leu Ser Ala 20 25 30 35	393
60	ATC CCC CAG CCG ACC ATC ACC AAG CAG GAA GTT ATC TAG CAA GCC GCT Ile Pro Gln Pro Thr Ile Thr Lys Gln Glu Val Ile Gln Ala Ala 40 45 50	441
65	GGG GCT TGG GGG CTC CAC TGG CTC CCC CCA GCC CCC TAA GAG AGC ACC Gly Ala Trp Gly Leu His Trp Leu Pro Pro Ala Pro Glu Ser Thr 55 60 65	489
70	TGG TGA TCA CGT GGT CAC GGC AAA GGA AGA CGT GAT GCC AGG ACC AGT Trp Ser Arg Gly His Gly Lys Gly Arg Arg Asp Ala Arg Thr Ser 70 75 80	537
75	CCC AGA GCA GGA ATG GGA AGG ATG AAG GGC CCG AGA ACA TGG CCT AAG Pro Arg Ala Gly Met Gly Arg Met Lys Gly Pro Arg Thr Trp Pro Lys 85 90 95	585
80	GCA CAT CCC ACT GCA CCC TGA CGC CCT GCT CTG ATA ACA AGA CTT Ala His Pro Thr Ala Pro Arg Pro Ala Leu Ile Thr Arg Leu 100 105 110	630
85	TGA CTT GGG GAG ACC CTC TAC TGC CTT GGA CAA CTT TCT CAT GTT GAA	678

Leu Gly Glu Thr Leu Tyr Cys Leu Gly Gln Leu Ser His Val Glu
 115 120 125

5 GCC ACT GCC TTC ACC TTC ACC TTC ATC CAT GTC CAA CCC CCG ACT TCA
 Ala Thr Ala Phe Thr Phe Thr Ile His Val Gln Pro Pro Thr Ser
 130 135 140

10 TCC CAA AGG ACA GCC GCC TGG AGA TGA CTT GAG CCT TAC TTA AAC CCA
 Ser Gln Arg Thr Ala Ala Trp Arg Leu Glu Pro Tyr Leu Asn Pro
 145 150 155

15 GCT CCC TTC TTC CCT AGC CTG GTG CTT CTC CTC TCC TAG CCC CGG TCA
 Ala Pro Phe Pro Ser Leu Val Leu Leu Ser Pro Arg Ser
 160 165 170

20 TGG TGT CCA GAC AGA GCC CTG TGA GGC TGG GTC CAA TTG TGG CAC TTG
 Trp Cys Pro Asp Arg Ala Leu Gly Trp Val Gln Leu Trp His Leu
 175 180 185

25 GGG CAC CTT GCT CCT TCT GCT GCT GCC CCC ACC TCT GCT GCC TCC
 Gly His Leu Ala Pro Pro Ser Ala Ala Ala Pro Thr Ser Ala Ala Ser
 190 195 200

30 CTC TGC TGT CAC CTT GCT CAG CCA TCC CGT CTT CTC CAA CAC CAC CTC
 Leu Cys Cys His Leu Ala Gln Pro Ser Arg Leu Leu Gln His His Leu
 205 210 215

35 TAC AGA GGC CAA GGA GGC CTT GGA AAC GAT TCC CCC AGT CAT TCT GGG
 Tyr Arg Gly Gln Gly Leu Gly Asn Asp Ser Pro Ser His Ser Gly
 220 225 230

40 AAC ATG TTG TAA GCA CTG ACT GGG ACC AGG CAC CAG GCA GGG TCT AGA
 Asn Met Leu Ala Leu Thr Gly Thr Arg His Gln Ala Gly Ser Arg
 235 240 245

45 AGG CTG TGG TGA GGG AAG ACG CCT TTC TCC TCC AAC CCA AC
 Arg Leu Trp Gly Lys Thr Pro Phe Ser Ser Asn Pro
 250 255 260

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 261 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

Ala Thr Pro Glu Thr Pro Gln Pro Ser Pro Pro Gly Gly Ser Gly Ser
 1 5 10 15

55 Glu Pro Tyr Lys Leu Leu Pro Gly Ala Val Ala Thr Ile Val Lys Pro
 20 25 30

Leu Ser Ala Ile Pro Gln Pro Thr Ile Thr Lys Gln Glu Val Ile Gln
 35 40 45

60 Ala Ala Gly Ala Trp Gly Leu His Trp Leu Pro Pro Ala Pro Glu Ser

50 55 60

Thr Trp Ser Arg Gly His Gly Lys Gly Arg Arg Asp Ala Arg Thr Ser
 65 70 75 80

5 Pro Arg Ala Gly Met Gly Arg Met Lys Gly Pro Arg Thr Trp Pro Lys
 85 90 95

Ala His Pro Thr Ala Pro Arg Pro Ala Leu Ile Thr Arg Leu Leu Gly
 10 100 105 - 110

Glu Thr Leu Tyr Cys Leu Gly Gln Leu Ser His Val Glu Ala Thr Ala
 115 120 125

15 Phe Thr Phe Thr Phe Ile His Val Gln Pro Pro Thr Ser Ser Gln Arg
 130 135 140

Thr Ala Ala Trp Arg Leu Glu Pro Tyr Leu Asn Pro Ala Pro Phe Phe
 145 150 155 160

20 Pro Ser Leu Val Leu Leu Leu Ser Pro Arg Ser Trp Cys Pro Asp Arg
 165 170 175

Ala Leu Gly Trp Val Gln Leu Trp His Leu Gly His Leu Ala Pro Pro
 25 180 185 190

Ser Ala Ala Ala Pro Thr Ser Ala Ala Ser Leu Cys Cys His Leu Ala
 195 200 205

30 Gln Pro Ser Arg Leu Leu Gln His His Leu Tyr Arg Gly Gln Gly Gly
 210 215 220

Leu Gly Asn Asp Ser Pro Ser His Ser Gly Asn Met Leu Ala Leu Thr
 35 225 230 235 240

Gly Thr Arg His Gln Ala Gly Ser Arg Arg Leu Trp Gly Lys Thr Pro
 245 250 255

40 Phe Ser Ser Asn Pro
 260

(2) INFORMATION FOR SEQ ID NO:56:

- 45 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

GGGCACGTGGG AGGAGGCAGT

20

55 (2) INFORMATION FOR SEQ ID NO:57:

- 60 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

5 GCCTGTAGGA CCAACCTACC

20

(2) INFORMATION FOR SEQ ID NO:58:

- 10 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

TCTGGTGTGC ACGACTGCAC

20

20 (2) INFORMATION FOR SEQ ID NO:59:

- 25 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

35 CTGGAGCTGC AGCCTCATAC

20

40 (2) INFORMATION FOR SEQ ID NO:60:

- 35 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

AAGGCTCCCT TAGATGCCCTG

20

50 (2) INFORMATION FOR SEQ ID NO:61:

- 50 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

CCACTCAGGG AGAAGACAGA CCT

23

(2) INFORMATION FOR SEQ ID NO:62:

- 5 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

CCTAGTTCTG TCCTAAAGAGG

20

15 (2) INFORMATION FOR SEQ ID NO:63:

- 20 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

GTCATAAAGT GTGGCTACAG

20

30 (2) INFORMATION FOR SEQ ID NO:64:

- 35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 22 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

45 CCACCCCCCTA CTCCATCCCT GT

22

50 (2) INFORMATION FOR SEQ ID NO:65:

- 55 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

60 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

50 CCCTCCCGTC AGCTGCTCCA

20

55 (2) INFORMATION FOR SEQ ID NO:66:

- 60 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

GTGCAGGGGA CAGAGAATGC

20

10

(2) INFORMATION FOR SEQ ID NO:67:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

AATCAAGCCA GTCCACGGCT AT

22

20

(2) INFORMATION FOR SEQ ID NO:68:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

GCCCAGCGTC ACTGAGTTGG CTA

23

30

(2) INFORMATION FOR SEQ ID NO:69:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

35 40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

TTGCCTGGGT GAGTGCCATG

20

45

(2) INFORMATION FOR SEQ ID NO:70:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

55 GCACCAGCTA TCTTGCCAAC

20

60

(2) INFORMATION FOR SEQ ID NO:71:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs

- (B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

AGGAGAAGTC TGGCAGAGCG

20

10 (2) INFORMATION FOR SEQ ID NO:72:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

20 CTCCTTGTGT GACACAAGTC

20

25 (2) INFORMATION FOR SEQ ID NO:73:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

35 CTCACTGTGT GAGGCCTGTC

20

40 (2) INFORMATION FOR SEQ ID NO:74:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

50 TGTTTGATTG GCCACGCCCTG

20

55 (2) INFORMATION FOR SEQ ID NO:75:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

60 ATCCTGGTTC TACCTTCTAG

20

(2) INFORMATION FOR SEQ ID NO:76:

- 5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

CATTTACTCC CACAAAGGCT

20

15 (2) INFORMATION FOR SEQ ID NO:77:

- 20 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

GACCACGTGA TCACCAGGTG

20

30 (2) INFORMATION FOR SEQ ID NO:78:

- 35 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1441 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

40 (ix) FEATURE:

- (A) NAME/KEY: CDS
(B) LOCATION: 20..1414

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

CTCCAAAACC CTCGTCGAC ATG GAC ATG GCC GAC TAC AGT GCT GCA CTG GAC
Met Asp Met Ala Asp Tyr Ser Ala Ala Leu Asp
1 5 10

52

50 CCA GCC TAC ACC ACC CTG GAA TTT GAG AAT GTG CAG GTG TTG ACG ATG
Pro Ala Tyr Thr Leu Glu Phe Glu Asn Val Gln Val Leu Thr Met
15 20 25

100

55 GGC AAT GAC ACG TCC CCA TCA GAA GGC ACC AAC CTC AAC GCG CCC AAC
Gly Asn Asp Thr Ser Pro Ser Glu Gly Thr Asn Leu Asn Ala Pro Asn
30 35 40

148

60 AGC CTG GGT GTC AGC GCC CTG TGT GCC ATC TGC GGG GAC CGG GCC ACG
Ser Leu Gly Val Ser Ala Leu Cys Ala Ile Cys Gly Asp Arg Ala Thr
45 50 55

196

65 GGC AAA CAC TAC GGT GCC TCG AGC TGT GAC GGC TGC AAG GGC TTC TTC
Gly Lys His Tyr Gly Ala Ser Ser Cys Asp Gly Cys Lys Gly Phe Phe
60 65 70 75

244

	CGG AGG AGC GTG CGG AAG AAC CAC ATG TAC TCC TGC AGA TTT AGC CGG Arg Arg Ser Val Arg Lys Asn His Met Tyr Ser Cys Arg Phe Ser Arg	80	85	90	292
5	CAG TGC GTG GTG GAC AAA GAC AAG AGG AAC CAG TGC CGC TAC TGC AGG Gln Cys Val Val Asp Lys Asp Lys Arg Asn Gln Cys Arg Tyr Cys Arg	95	100	105	340
10	CTC AAG AAA TGC TTC CGG GCT GCC ATG AAG AAG GAA GCC GTC CAG AAT Leu Lys Lys Cys Phe Arg Ala Gly Met Lys Lys Glu Ala Val Gln Asn	110	115	120	388
15	GAG CGG GAC CGG ATC AGC ACT CGA AGG TCA AGC TAT GAG GAC AGC AGC Glu Arg Asp Arg Ile Ser Thr Arg Arg Ser Ser Tyr Glu Asp Ser Ser	125	130	135	436
20	CTG CCC TCC ATC AAT GCG CTC CTG CAG GCG GAG GTC CTG TCC CGA CAG Leu Pro Ser Ile Asn Ala Leu Leu Gln Ala Glu Val Leu Ser Arg Gln	140	145	150	484
25	ATC ACC TCC CCC GTC TCC GGG ATC AAC GGC GAC ATT CGG GCG AAG AAG Ile Thr Ser Pro Val Ser Gly Ile Asn Gly Asp Ile Arg Ala Lys Lys	160	165	170	532
30	ATT GCC AGC ATC GCA GAT GTG TGT GAG TCC ATG AAG GAG CAG CTG CTG Ile Ala Ser Ile Ala Asp Val Cys Glu Ser Met Lys Glu Gln Leu Leu	175	180	185	580
35	GTT CTC GTT GAG TGG GCC AAG TAC ATC CCA GCT TTC TGC GAG CTC CCC Val Leu Val Glu Trp Ala Lys Tyr Ile Pro Ala Phe Cys Glu Leu Pro	190	195	200	628
40	CTG GAC GAC CAG GTG GCC CTG CTC AGA GCC CAT GCT GGC GAG CAC CTG Leu Asp Asp Gln Val Ala Leu Leu Arg Ala His Ala Gly Glu His Leu	205	210	215	676
45	CTG CTC GGA GCC ACC AAG AGA TCC ATG GTG TTC AAG GAC GTG CTG CTC Leu Leu Gly Ala Thr Lys Arg Ser Met Val Phe Lys Asp Val Leu Leu	220	225	230	724
50	CTA GGC AAT GAC TAC ATT GTC CCT CGG CAC TGC CCG GAG CTG GCG GAG Leu Gly Asn Asp Tyr Ile Val Pro Arg His Cys Pro Glu Leu Ala Glu	240	245	250	- 772
55	ATG AGC CGG GTG TCC ATA CGC ATC CTT GAC GAG CTG GTG CTG CCC TTC Met Ser Arg Val Ser Ile Arg Ile Leu Asp Glu Leu Val Leu Pro Phe	255	260	265	820
60	CAG GAG CTG CAG ATC GAT GAC AAT GAG TAT GCC TAC CTC AAA GCC ATC Gln Glu Leu Gln Ile Asp Asp Asn Glu Tyr Ala Tyr Leu Lys Ala Ile	270	275	280	868
	ATC TTC TTT GAC CCA GAT GCC AAG GGG CTG AGC GAT CCA GGG AAG ATC Ile Phe Phe Asp Pro Asp Ala Lys Gly Leu Ser Asp Pro Gly Lys Ile	285	290	295	916
	AAG CGG CTG CGT TCC CAG GTG CAG GTG AGC TTG GAG GAC TAC ATC AAC Lys Arg Leu Arg Ser Gln Val Gln Val Ser Leu Glu Asp Tyr Ile Asn	300	305	310	964
				315	

	GAC CGC CAG TAT GAC TCG CGT GGC CGC TTT GGA GAG CTG CTG CTG CTG Asp Arg Gln Tyr Asp Ser Arg Gly Arg Phe Gly Glu Leu Leu Leu 320 325 330	1012
5	CTG CCC ACC TTG CAG AGC ATC ACC TGG CAG ATG ATC GAG CAG ATC CAG Leu Pro Thr Leu Gln Ser Ile Thr Trp Gln Met Ile Glu Gln Ile Gln 335 340 345	1060
10	TTC ATC AAG CTC TTC GGC ATG GCC AAG ATT GAC AAC CTG TTG CAG GAG Phe Ile Lys Leu Phe Gly Met Ala Lys Ile Asp Asn Leu Leu Gln Glu 350 355 360	1108
15	ATG CTG CTG GGA GGG TCC CCC AGC GAT GCA CCC CAT GCC CAC CAC CCC Met Leu Leu Gly Gly Ser Pro Ser Asp Ala Pro His Ala His His Pro 365 370 375	1156
20	CTG CAC CCT CAC CTG ATG CAG GAA CAT ATG GGA ACC AAC GTC ATC GTT Leu His Pro His Leu Met Gln Glu His Met Gly Thr Asn Val Ile Val 380 385 390 395	1204
25	GCC AAC ACA ATG CCC ACT CAC CTC AGC AAC GGA CAG ATG TGT GAG TGG Ala Asn Thr Met Pro Thr His Leu Ser Asn Gly Gln Met Cys Glu Trp 400 405 410	1252
30	CCC CGA CCC AGG GGA CAG GCA GCC ACC CCT GAG ACC CCA CAG CCC TCA Pro Arg Pro Arg Gly Gln Ala Ala Thr Pro Glu Thr Pro Gln Pro Ser 415 420 425	1300
35	CCG CCA GGT GCG TCA GGG TCT GAG CCC TAT AAG CTC CTG CCG GGA GCC Pro Pro Gly Ala Ser Gly Ser Glu Pro Tyr Lys Leu Leu Pro Gly Ala 430 435 440	1348
40	GTC GCC ACA ATC GTC AAG CCC CTC TCT GCC ATC CCC CAG CCG ACC ATC Val Ala Thr Ile Val Lys Pro Leu Ser Ala Ile Pro Gln Pro Thr Ile 445 450 455	1396
	ACC AAG CAG GAA GTT ATC TAGCAAGCCG CTGGGGCTTG GGGGCTC Thr Lys Gln Glu Val Ile 460 465	1441

(2) INFORMATION FOR SEQ ID NO:79:

45 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 465 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

50 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:

Met	Asp	Met	Ala	Asp	Tyr	Ser	Ala	Ala	Leu	Asp	Pro	Ala	Tyr	Thr	Thr	
1		5					10					15				
55	Leu	Glu	Phe	Glu	Asn	Val	Gln	Val	Leu	Thr	Met	Gly	Asn	Asp	Thr	Ser
	20					25						30				
60	Pro	Ser	Glu	Gly	Thr	Asn	Leu	Asn	Ala	Pro	Asn	Ser	Leu	Gly	Val	Ser
		35				40						45				

	Ala Leu Cys Ala Ile Cys Gly Asp Arg Ala Thr Gly Lys His Tyr Gly			
	50	55	60	
5	Ala Ser Ser Cys Asp Gly Cys Lys Gly Phe Phe Arg Arg Ser Val Arg			
	65	70	75	80
	Lys Asn His Met Tyr Ser Cys Arg Phe Ser Arg Gln Cys Val Val Asp			
	85	90	95	
10	Lys Asp Lys Arg Asn Gln Cys Arg Tyr Cys Arg Leu Lys Lys Cys Phe			
	100	105	110	
	Arg Ala Gly Met Lys Lys Glu Ala Val Gln Asn Glu Arg Asp Arg Ile			
15	115	120	125	
	Ser Thr Arg Arg Ser Ser Tyr Glu Asp Ser Ser Leu Pro Ser Ile Asn			
	130	135	140	
20	Ala Leu Leu Gln Ala Glu Val Leu Ser Arg Gln Ile Thr Ser Pro Val			
	145	150	155	160
	Ser Gly Ile Asn Gly Asp Ile Arg Ala Lys Lys Ile Ala Ser Ile Ala			
	165	170	175	
25	Asp Val Cys Glu Ser Met Lys Glu Gln Leu Leu Val Leu Val Glu Trp			
	180	185	190	
	Ala Lys Tyr Ile Pro Ala Phe Cys Glu Leu Pro Leu Asp Asp Gln Val			
30	195	200	205	
	Ala Leu Leu Arg Ala His Ala Gly Glu His Leu Leu Leu Gly Ala Thr			
	210	215	220	
35	Lys Arg Ser Met Val Phe Lys Asp Val Leu Leu Leu Gly Asn Asp Tyr			
	225	230	235	240
	Ile Val Pro Arg His Cys Pro Glu Leu Ala Glu Met Ser Arg Val Ser			
40	245	250	255	
	Ile Arg Ile Leu Asp Glu Leu Val Leu Pro Phe Gln Glu Leu Gln Ile			
	260	265	270	
	Asp Asp Asn Glu Tyr Ala Tyr Leu Lys Ala Ile Ile Phe Phe Asp Pro			
45	275	280	285	
	Asp Ala Lys Gly Leu Ser Asp Pro Gly Lys Ile Lys Arg Leu Arg Ser			
	290	295	300	
50	Gln Val Gln Val Ser Leu Glu Asp Tyr Ile Asn Asp Arg Gln Tyr Asp			
	305	310	315	320
	Ser Arg Gly Arg Phe Gly Glu Leu Leu Leu Leu Pro Thr Leu Gln			
	325	330	335	
55	Ser Ile Thr Trp Gln Met Ile Glu Gln Ile Gln Phe Ile Lys Leu Phe			
	340	345	350	
	Gly Met Ala Lys Ile Asp Asn Leu Leu Gln Glu Met Leu Leu Gly Gly			
60	355	360	365	
	Ser Pro Ser Asp Ala Pro His Ala His His Pro Leu His Pro His Leu			

	370	375	380	
	Met Gln Glu His Met Gly Thr Asn Val Ile Val Ala Asn Thr Met Pro			
5	385	390	395	400
	Thr His Leu Ser Asn Gly Gln Met Cys Glu Trp Pro Arg Pro Arg Gly			
	405	410	415	
10	Gln Ala Ala Thr Pro Glu Thr Pro Gln Pro Ser Pro Pro Gly Ala Ser			
	420	425	430	
	Gly Ser Glu Pro Tyr Lys Leu Leu Pro Gly Ala Val Ala Thr Ile Val			
	435	440	445	
15	Lys Pro Leu Ser Ala Ile Pro Gln Pro Thr Ile Thr Lys Gln Glu Val			
	450	455	460	
	Ile			
20	465			
	(2) INFORMATION FOR SEQ ID NO:80:			
25	(i) SEQUENCE CHARACTERISTICS:			
	(A) LENGTH: 2329 base pairs			
	(B) TYPE: nucleic acid			
	(C) STRANDEDNESS: single			
	(D) TOPOLOGY: linear			
30	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:			
	GGGGCCCTGA TTCACGGGCC GCTGGGGCAG CGTTGGGGGT TGGGGGTGCC CACAGGGTTG		60	
35	GCTAGTGGGG TTTTGGGGGG GCAGTGCGTG CAAGGAGTTT GGTTTGTGTC TGCCGGCCGG		120	
	CAGGCACAAACG CAACCACGCG GTGGGGGAGG CGGCTAGCGT GGTGGACGGC CCGCGTGGCC		180	
40	CTGTGGCAGC CGAGCCATGG TTTCTAAACT GAGCCAGCTG CAGACGGAGC TCCTGGCGGC		240	
	CCTGCTCGAG TCAGGGCTGA GCAAAGAGGC ACTGATCCAG GCACTGGGTG AGCCGGGCC		300	
	CTACCTCCTG GCTGGAGAAG GCCCCCTGGA CAAGGGGGAG TCCTGCGGCC GCGGTCGAGG		360	
45	GGAGCTGGCT GAGCTGCCA ATGGGCTGGG GGAGACTCGG GGCTCCGAGG ACGAGACGGA		420	
	CGACGATGGG GAAGACTTCA CGCCACCCAT CCTCAAAGAG CTGGAGAACCC TCAGCCCTGA		480	
	GGAGGCGGCC CACCAAGAAG CGTGGTGGGA GACCCTCTG CAGGAGGACC CGTGGCGTGT		540	
50	GGCGAAGATG GTCAAGTCCT ACCTGCAGCA GCACAACATC CCACAGCGGG AGGTGGTCGA		600	
	TACCACTGGC CTCAACCAGT CCCACCTGTC CCAACACCTC AACAAAGGGCA CTCCCCTGAA		660	
	GACGCAGAAG CGGGCCGCC TGTACACCTG GTACGTCCGC AAGCAGCGAG AGGTGGCGCA		720	
55	GCAGTTCACC CATGCAGGGC AGGGAGGGCT GATTGAAGAG CCCACAGGTG ATGAGCTACC		780	
	AACCAAGAAG GGGCGGAGGA ACCGTTCAA GTGGGGCCCA GCATCCCAGC AGATCCTGTT		840	
60	CCAGGCCTAT GAGAGGCAGA AGAACCCCTAG CAAGGAGGAG CGAGAGACGC TAGTGGAGGA		900	

Page No. 30 of 30

	GTGCAATAGG GCGGAATGCA TCCAGAGAGG GGTGTCCCCA TCACAGGCAC AGGGGCTGGG	960
	CTCCAACCTC GTCACGGAGG TGCGTGTCTA CAACTGGTT GCCAACCAGC GCAAAGAAGA	1020
5	AGCCTTCCGG CACAAGCTGG CCATGGACAC GTACAGCGGG CCCCCCCCAG GGCCAGGCC	1080
	GGGACCTGCG CTGCCCCTC ACAGCTCCCC TGGCCTGCCT CCACCTGCC C TCTCCCCAG	1140
10	TAAGGTCCAC GGTGTGCGCT ATGGACAGCC TGCGACCAGT GAGACTGCAG AAGTACCCCTC	1200
	AAGCAGCGGC GGTCCCTTAG TGACAGTGTC TACACCCCTC CACCAAGTGT CCCCCACGGG	1260
	CCTGGAGCCC AGCCACAGCC TGCTGAGTAC AGAACCAAG CTGGTCTCAG CAGCTGGGG	1320
15	CCCCCTCCCC CCTGTCAGCA CCCTGACAGC ACTGCACAGC TTGGAGCAGA CATCCCCAGG	1380
	CCTCAACCAG CAGCCCCAGA ACCTCATCAT GCCCTCACTT CCTGGGGTCA TGACCATCGG	1440
20	GCCTGGTGAG CCTGCCTCCC TGGGTCTAC GTTCACCAAC ACAGGTGCCT CCACCTGGT	1500
	CATCGGCCTG GCCTCCACGC AGGCACAGAG TGTGCCGGTC ATCAACAGCA TGGGCAGCAG	1560
	CCTGACCACC CTGCAGCCCG TCCAGTTCTC CCAGCCGCTG CACCCCTCCT ACCAGCAGCC	1620
25	GCTCATGCCA CCTGTGCAGA GCCATGTGAC CCAGAGCCCC TTCATGGCCA CCATGGCTCA	1680
	GCTGCAGAGC CCCCACGGTG AGCACCCCTGT GCCCCACACA GCAGGAGATG ATGATAGAGG	1740
	TTGGCTGTCA ATGGATGCAG GGGAAAGGGG TGCCTGGCAG GCATTGCAGT CTGCATGTGT	1800
30	CTCTGGGACA AGTGTTTTC CGTGATTGAG GGTGTCTGCA GGCCAGTGTG TTCCCATGTG	1860
	AATGCACGTA TCTGTGTGTG TGCACGACTG CTTGTGTGAG CAGATCCCTA GTCGTGTCTG	1920
35	GGTGTGTATC GGTTGTGCAT GCATTGTGT GCATCCTGTG TTTCTCTGAA ACTCTTAGGG	1980
	CCATATGAAT TTCTAAAATC TATTCAGATT TTAGAAAGGT AATCTGGGC CAGGCAGGGT	2040
	GGCTCATGCC TGTAATCCCA GCACTTTGGA AGGCCGAGGT GGGCAGATCA CTTGAGGTCA	2100
40	GGAGTTCAAG ACCAGCCTGG CCAACACGGT GAAACCCGT CTCTACTAAA AGTACAAAAAA	2160
	TTAGCCAGGC GTGGAGCACG TGCCTGTAGT CCCAGCTACT TGGGAGGCTG AGGCAGAAC	2220
45	GCTTGAACCT GGGAGGCGGA GGTTGCAGTG AGCTGAGATT TGGCCACTGC ACTGCACTCC	2280
	AGCCTGGGCA ACAGAGTGAG TACTCTGCCA AAAAAAAAAA AAAAAAAAAA	2329

50 (2) INFORMATION FOR SEQ ID NO:81:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:

60 CACCTGGTGA TCACGTGGTC 20

(2) INFORMATION FOR SEQ ID NO:82:

- 5 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

GTAAGGCTCA AGTCATCTCC

20

15 (2) INFORMATION FOR SEQ ID NO:83:

- 20 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 5 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

Glu Gly Cys Lys Gly
1 5

30 (2) INFORMATION FOR SEQ ID NO:84:

- 35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 5 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

Glu Gly Cys Lys Ala
1 5

45 (2) INFORMATION FOR SEQ ID NO:85:

- 50 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 5 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:

Asp Gly Cys Lys Gly
1 5

60 (2) INFORMATION FOR SEQ ID NO:86:

- 65 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 36 base pairs
 (B) TYPE: nucleic acid

- (C) STRANDEDNESS: single
(D) TOPOLOGY: linear

5 (ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..36

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:

GAC ACG TAC AGC GGC CCC CCC CCA GGG CCA GGC CCG
Asp Thr Tyr Ser Gly Pro Pro Pro Gly Pro Gly Pro
1 5 10

36

15 (2) INFORMATION FOR SEQ ID NO:87:
20 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

25 (ii) MOLECULE TYPE: protein

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:

Asp Thr Tyr Ser Gly Pro Pro Pro Gly Pro Gly Pro
1 5 10

35 (2) INFORMATION FOR SEQ ID NO:88:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

40 (ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..36

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:

GAC ACG TAC AGC GGC CCC CCC AGG GCC AGG CCC
Asp Thr Tyr Ser Gly Pro Pro Pro Arg Ala Arg Pro
1 5 10

36

55 (2) INFORMATION FOR SEQ ID NO:89:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

60 (ii) MOLECULE TYPE: protein

5
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:

Asp Thr Tyr Ser Gly Pro Pro Pro Arg Ala Arg Pro
1 5 10

10
(2) INFORMATION FOR SEQ ID NO:90:

- 15
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

20
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:

CATGAACCCC GAAGAGTGTT G

21

25
(2) INFORMATION FOR SEQ ID NO:91:

- 30
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

35
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:91:

GCCTCCAGAC ACCTGTTACT

20

40
(2) INFORMATION FOR SEQ ID NO:92:

- 45
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

50
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:92:

GGCGATCATG GCAAGTTAGA AG

22

55
(2) INFORMATION FOR SEQ ID NO:93:

- 55
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

TTGGTGAGAG TATGGAAGAC C

21

(2) INFORMATION FOR SEQ ID NO:94:

- 5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:94:

GGGGTTTGCT TGTGAACTC C

21

15 (2) INFORMATION FOR SEQ ID NO:95:

- 20 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:95:

TTGGTGGGAA ACGGGCTTGG

20

30 (2) INFORMATION FOR SEQ ID NO:96:

- 35 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:96:

CTCCCCACTAG TACCCTAACCC

20

45 (2) INFORMATION FOR SEQ ID NO:97:

- 50 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:97:

50 GAGAGGGCAA AGGTCACTTC AG

22

60 (2) INFORMATION FOR SEQ ID NO:98:

- 55 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:98:

AGTGAAGGCT ACAGACCCTA TC

22

5

(2) INFORMATION FOR SEQ ID NO:99:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:99:

10

TTCCTGGGTC TGTGTACTTG C

21

15

(2) INFORMATION FOR SEQ ID NO:100:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:100:

TGTGTTTG GCCAAGCACC A

21

20

(2) INFORMATION FOR SEQ ID NO:101:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:101:

AACCAGATAA GATCCGTGGC

20

30

(2) INFORMATION FOR SEQ ID NO:102:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:102:

40

AACCAGACTC ACAGCCTGAA CC

22

45

(2) INFORMATION FOR SEQ ID NO:103:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs

- (B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:103:

TCACAGGGCA ATGGCTGAAC

20

10 (2) INFORMATION FOR SEQ ID NO:104:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:104:

20 TGCCGAGTCA TTGTTCCAGG

20

25 (2) INFORMATION FOR SEQ ID NO:105:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:105:

35 CCTCTTATCT TATCAGCTCC AG

22

40 (2) INFORMATION FOR SEQ ID NO:106:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:106:

50 CTGCTCTTG TGGTCCAAGT CC

22

55 (2) INFORMATION FOR SEQ ID NO:107:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

60 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:107:

GAGTTTGAAG GAGACCTACA G

21

(2) INFORMATION FOR SEQ ID NO:108:

- 5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:108:

ATCCACCTCT CCTTATCCCC G

21

15 (2) INFORMATION FOR SEQ ID NO:109:

- 20 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:109:

ACTTCCGAGA AAGTTCA GAC C

21

30 (2) INFORMATION FOR SEQ ID NO:110:

- 35 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:110:

TTTGCCTGTG TATGCACCTT G

21

45 (2) INFORMATION FOR SEQ ID NO:111:

- 50 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:111:

55 GCCGAGTCCA TGCTTGCCAC

20

(2) INFORMATION FOR SEQ ID NO:112:

- 60 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:112:

CTTTGCTGGT TGAGTTGGGC

20

5 (2) INFORMATION FOR SEQ ID NO:113:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
10 (B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:113:

15 TTCCATGACA GCTGCCAGA G

21

20 (2) INFORMATION FOR SEQ ID NO:114:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
25 (B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:114:

30 TAAAGGTTGG AGCCCCTCTG

20

35 (2) INFORMATION FOR SEQ ID NO:115:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
40 (B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:115:

45 TTGTAAGGTG ACCCCATCAG

20

50 (2) INFORMATION FOR SEQ ID NO:116:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
55 (B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:116:

60 TTGGTGATGT CCAGAAAGTCC

20

(2) INFORMATION FOR SEQ ID NO:117:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
65 (B) TYPE: nucleic acid

- (C) STRANDEDNESS: single
(D) TOPOLOGY: linear

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:117:

CAGAATGTGT CAGAGTCGC

20

10 (2) INFORMATION FOR SEQ ID NO:118:

- 15 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:118:

20 CTCCCTCCTG TTCTTAAGTG

20

25 (2) INFORMATION FOR SEQ ID NO:119:

- 30 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:119:

35 CTGGACTCCC AGTCAGTCA

20

40 (2) INFORMATION FOR SEQ ID NO:120:

- 45 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:120:

50 CAAGGATCCA GAAGATTGGC

20

55 (2) INFORMATION FOR SEQ ID NO:121:

- 60 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:121:

CGTCCTCTGG GAAGATCTGC

20

5 (2) INFORMATION FOR SEQ ID NO:122:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:122:

GCAACAGAGC AAGACTCCAT CTCA

24

15 (2) INFORMATION FOR SEQ ID NO:123:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
20 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:123:

GAGTTTAATG GAAGAACTAA CC

22

25 (2) INFORMATION FOR SEQ ID NO:124:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:124:

CCTCATGGAG AAACATCCTA AGT

23

35 (2) INFORMATION FOR SEQ ID NO:125:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
40 (C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:125:

45 50 AGGGAGTGCA CGGCTGAGCT CCTG

24

55 (2) INFORMATION FOR SEQ ID NO:126:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6254 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
60 (D) TOPOLOGY: linear

5 (ix) FEATURE:

- (A) NAME/KEY: modified_base
(B) LOCATION: 1287..4273
(D) OTHER INFORMATION: /note= "N = A or G or C or T"

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:126:

AGCCAGCACT	GTTCTGGCA	CATGGTAATC	TTAACATATT	TTTCCTACA	GGGAGGCCTG	60
GTGTCAGGCC	GGGAGTGGGG	TGGAAGGGTC	CCAAAATGGA	TGGAAGGGCC	CCAAAATGGC	120
CGTGAGCATC	CTCTGCCCTT	GAGAAGAGCT	AGCCCAGCTG	TCTAGAGCTC	CCTGCTGCTG	180
15 CCGCTCTCGT	AAGCAGCAAG	CATTTTGCG	TCTCCTGTCT	CAGCATGATG	CCCTACAAAG	240
GTTCTTCGG	GGGTGGGACC	CAACGCTGCT	CTCCTGATGG	CCTCCCTGGC	TCCCAGCACC	300
20 TTCCATCCCA	GCTGCTCAGG	GCCCCTCAC	TGCGCCTCCC	CCACCCCTCCC	CTCTGCCAC	360
TCCCCATCGCA	GGCCATAGCT	CCCTGTCCCT	CTCCGCTGCC	ATGAGGCCTG	CACTTGCAG	420
25 GGCTGAAGTC	CAAAGTTCAAG	TCCCTTCGCT	AAGCACACGG	ATAAAATATGA	ACCTTGGAGA	480
ATTTCAGGAG	CTCCAATGTA	AACAGAACAG	GCAGGGGCC	TGATTCACGG	GCCGCTGGGG	540
30 CCAGGGTTGG	GGGTTGGGG	TGCCCACAGG	GCTTGGCTAG	TGGGGTTTTG	GGGGGGCAGT	600
GGGTGCAAGG	AGTTTGGTT	GTGTCTGCCG	GCCGGCAGGC	AAACGCAACC	CACGCCGTGG	660
35 GGGAGGCCGC	TAGCGTGGTG	GACCCGGGCC	CGTGGCCCT	GTGGCAGCCG	AGCCATGGTT	720
TCTAAACTGA	GCCAGCTGCA	GACGGAGCTC	CTGGCGGCC	TGCTCGAGTC	AGGGCTGAGC	780
40 AAAGAGGCAC	TGATCCAGGC	ACTGGGTGAG	CCGGGGCCCT	ACCTCCTGGC	TGGAGAAGGC	840
CCCCTGGACA	AGGGGGAGTC	CTGCGCGGC	GGTCGAGGGG	AGCTGGCTGA	GCTGCCAAT	900
45 GGGCTGGGGG	AGACTCGGGG	CTCCGAGGAC	GAGACGGACG	ACGATGGGA	AGACTTCACG	960
CCACCCATCC	TCAAAGAGCT	GGAGAACCTC	AGCCCTGAGG	AGGCGGCCA	CCAGAAAGCC	1020
GTGGTGGAGA	CCCTTCTGCA	GTAAGGAGCC	CTGCCCCGTC	CCCGCTCCCA	GGAGAGCCTA	1080
50 GAGGGGCC	CCTCAGCTCC	TAACGAGCCC	CCCTTCTGAG	TTGAGTCCCC	ATGACCTTCA	1140
GCCTTAGCC	TAGTTGCTGG	GAAGGGGGAC	AGGGCCCATG	AGAGCCCAGG	GGTCCTTGCT	1200
TGGAGGTTTG	AGCCTCCAGC	CCCTGAAGTC	CTCCTCTGCA	GAGTCCAAA	TCCCATGAGC	1260
55 CCAGGCCTT	AGCCCAGTCC	TTGGGCNAGG	GGGACATTTC	CCAGGGGGTC	CAAGATGGGA	1320
GAAAAAGCAG	TGAATTACA	ACTCAAATGC	CCACCCACCC	ATCCATCCAT	CCGTCCATCC	1380
ACCCATTCCAT	CCATTCCATCC	ATTCAACCCAT	CCATCCATCC	ACATATCTTC	ATCTGTGTTG	1440
60 TGTGTCTGTG	TATCCATGTT	TCTAACCTT	TATCTGTTCC	AGTGTCTGTA	TCCATAGGCC	1500
TGTGTCCACG	TTTGTCAATGT	GTGTGCGTCN	ACAAGTCTCT	GTCCTCATGA	CCATGTGTCT	1560
GTGTCCCTGT	GTCCTGGCAT	AAATGACCAT	ACCTCACCGT	CCCTGAGTCT	ATGTGTAGGC	1620

	CCCTGGGCTC CATAACTGCT TTCATGCACA GTCCCCACCC TCAGAGTTGA CAAGGTTCCA	1680
5	GCACCCAGGA CCGCAGCCCC ACCTATGGGG AGAGACAGCC CTTGCTGAGC AGATCCCGTC	1740
	CTTGCCTCT CCCAGGGAGG ACCCGTGGCG TGTGGCGAAG ATGGTCAAGT CCTACCTGCA	1800
	GCAGCACAAAC ATCCCACAGC GGGAGGTGGT CGATACCACT GCCCTCAACC AGTCCCACCT	1860
10	GTCCCAACAC CTCAACAAGG GCACCTCCAT GAAGACGCAG AAGCGGGCCG CCCTGTACAC	1920
	CTGGTACGTC CGCAAGCAGC GAGAGGTGGC GCAGCGTAAG TAATGACCCT ACCCCGCATC	1980
15	TTCCCTGGGA GGGCCCAGGA CTCTCCCCTA ACTCATAGGT GGGGGCTGGA AGCTTCACCA	2040
	TCCCCATTAC ACAGACAGGT AGATGGAAAG GAAGTCAGTG GGATTCAACC TGCATTTATT	2100
	ACCTATTCTG CGCCAGGCAC TCTGTGGGAC GGGAGTANAC TTGGTCCTGA ACATCCAAAG	2160
20	ATGAATGAAA TGGGTCCCTG CTTTCTTTTT CTTTTTTTAG ATACGTGACT CTGGAAAAAT	2220
	ATGTAAGCTC TCTGAGCCTC AGCTTCTTCA TCTGTACAAT GGGGATAGTA AATGTGCCAA	2280
25	ATCAGAACAA ATGCTAATGC TTACCTGCAG TCTTGTACTG AGAAGGGATGG TGAGATCATA	2340
	TCTTGGGTTG GTAGGAAAGC ATTCAAGGGAT TGATTAGTGA TGTTTGCCTT GAACACAGGT	2400
	TAAGAAAGTG ATGGCATGTG TGCTGTGTGT TTGTCATCAG TAGATTAGAT GATTCTAAG	2460
30	TTCTAGCTGT AAGCTCCTCT GGTCAGCGC CATGGCAATG AGAAAGAAC AAGGGCAAGG	2520
	TCAGGGGAAT GGACGAGGGA AGGTGAGAGT GGCCAGTACC CCACTCACGG CTTTCTGTGC	2580
35	CTGCAGAGTT CACCCATGCA GGGCAGGGAG GGCTGATTGA AGAGCCCACA GGTGATGAGC	2640
	TACCAACCAA GAAGGGCGG AGGAACCGTT TCAAGTGGGG CCCAGCATCC CAGCAGATCC	2700
	TGTTCCAGGC CTATGAGAGG CAGAAGAACCC CTAGCAAGGA GGAGCGAGAG GTACAACGGC	2760
40	GGGCGGGAAA CAGTGCTGGT TTGGTCTGGG CTGCGGCAAG GCCAGGGGAA GGGGAAGGTG	2820
	ACTCTAGGTC CTGTAAAAGG CTGTCCAGTT GCCGAGAACT CCTGATATTG GCTTAGCCTG	2880
45	GCCCAGAAAA TTGAGAATAC TTGAACCTAA GCCCATTCTCT CGCAGCCCCC CTGCACCNTG	2940
	GACACCAAGC AACCCCTTCC ATGGATGCTC ACCCAATTG ATTCTCTCTA CAATCCTATG	3000
	GCTCTTTGC TCACTTTATG AATGGAGAGA CTGAGGTCAAG ACAGACTGTC AATTGCCAA	3060
50	GGTCACACAG CAGACCTGGC ATTGGAACCC AGATCTGCCA GCCTCAAACC CTCCGGCAGA	3120
	GNTCAGCTTC TCAGAACCCCT CCCCTTCATG CCCAGGACAG GGTCCTCTG AGCCTGGCCT	3180
55	GGAGGCTCAT GGGTGGCTAT TTCTGCAGGG CGGAATGCAT CCAGAGAGGG GTGTCCCCAT	3240
	CACAGGCACA GGGGCTGGGC TCCAACCTCG TCACGGAGGT GCGTGTCTAC AACTGGTTTG	3300
	CCAACCGGCG CAAAGAAGAA GCCTTCCGGC ACAAGCTGGC CATGGACACG TACAGCGGGC	3360
60	CCCCCCAGG GCCAGGCCCG GGACCTGCGC TGCCCCCTCA CAGCTCCCCCT GGCCTGCCTC	3420

	CACCTGCCCT	CTCCCCAGT	AAGGTCCACG	GTAAGTGGTA	TGTGGGACA	AGGGACACGT	3480
	GGGAAGGTGG	GAGGGTTGGG	GAGGACTGTC	CCATTGACAG	CAGTCACCTA	AACCTCTTG	3540
5	CACGTCAGTT	TGGTTCCATT	CGCAGCTGAC	CCAGGGATTG	GCAAAAGGTA	GAAACAAAGG	3600
	CAGATTGCT	GGCTGCATAA	AGGCAGACAG	GCAGATGGCC	TAAGCAAACC	AATGGAGTTT	3660
10	GAAGTGCTGA	GGGCTGTGGA	GGCAGGGGAG	GGCAGGGAAG	TGGGGTGCTG	AGGCAGGACA	3720
	CTGCTTCCCT	CTCCAGGTGT	GCGCTATGGA	CAGCCTGCGA	CCAGTGAGAC	TGCAGAAGTA	3780
	CCCTCAAGCA	CGGGCGGTCC	CTTAGTGACA	GTGTCTACAC	CCCTCCACCA	AGTGTCCCCC	3840
15	ACGGGCCTGG	AGCCCAGCCA	CAGCCTGCTG	AGTACAGAAG	CCAAGCTGGT	GAGTGTCCCTT	3900
	GCTTGTAAGG	AAAACCAAC	CTCATTTTC	CTTGGCAGGG	AGATTCTGGA	GCAGTCCCTA	3960
20	GGGAGGCCCT	GTGGGGACCC	CGGCCCCCG	GACACAGCTT	GGCTTCCCT	CGTAGGTCTC	4020
	AGCAGCTGGG	GGCCCCCTCC	CCCCTGTCAG	CACCTGACA	GCACTGCACA	GCTTGGAGCA	4080
25	GACATCCCCA	GGCCTCAACC	AGCAGCCCCA	GAACCTCATC	ATGGCCTCAC	TTCCTGGGTT	4140
	CATGACCATC	GGGCCTGGTG	AGCCTGCCTC	CCTGGGTCCCT	ACGTTCACCA	ACACAGGTGC	4200
	CTCCACCTG	GTCATCGGTA	AGCTGGTGGG	GATGGGTGGG	CACCTGGGTG	GGAGGCTCAT	4260
30	GGGGCAACCG	CANAATCCAG	GAGCTGGAAA	AGCCACTGGG	ACTCATTATC	TCATTCAATTC	4320
	ATTCATACAA	CATGTTAGGA	GAGGGGAGCA	GAGAACTGAC	CCCATGGCCT	TTGCACTGCT	4380
	GTGGTACCCC	AGGGCTCCAG	GGAACCGCAG	TTTGACAAC	TTTGAACAAG	TCACCGCTTG	4440
35	CTTTTCCCAT	TAGCTTAGAC	AAAGAGCTAA	AGGCTCAGAG	AGGGGAAATG	ACTTGCCAGA	4500
	GCCACTTAAA	TTAGTGGCAG	GTCCCAGTGG	AGGGCTGTTT	CCTGACCACC	TTGCCCTTC	4560
	TTCCAAACCA	CGGGCTCTGG	GAAGGAGAGG	TGGTGCCCTT	GGGAGGTCTT	GGGCAGGGGT	4620
40	GGGATATAAC	TGGGGGCC	AGCTGATTCC	CTCCCCCTCC	ACTCCAGGCC	TGGCCTCCAC	4680
	GCAGGCACAG	AGTGTGCCGG	TCATCAACAG	CATGGCAGC	AGCCTGACCA	CCCTGCAGCC	4740
45	CGTCCAGTTC	TCCCAGCCGC	TGCACCCCTC	CTACCAGCAG	CCGCTCATGC	CACCTGTGCA	4800
	GAGCCATGTG	ACCCAGAAC	CCTTCATGGC	CACCATGGCT	CAGCTGCAGA	GCCCCCACGG	4860
	TGAGCACCC	GTGCCACCA	CAGCAGGAGA	TGATGATAGA	GGTTGGCTGT	CAATGGATGC	4920
50	AGGGGAAAGG	GGTGCCTGGC	AGGCATTGCA	GTCTGCATGT	GTCTCTGGGA	CAAGTGTGTT	4980
	TCCGTGATTG	AGGGTGTCTG	CAGGCCAGTG	TGTTCCCATG	TGAATGCACG	TATCTGTGTG	5040
55	TGTGCACGAC	TGCTTGTGTG	AGCAGATCCC	TAGTGCAGTGT	CTGGGTGTGT	ATCGGTTGTG	5100
	CATGCATTG	TGTGCATGCC	TGTGTTCTC	TGAAACTCTT	AGGGCCATAT	GAATTCTAA	5160
60	AATCTATTCA	GACCAGTTT	GAAAATCAGC	CTTGGATCTC	CAACTGCTGC	CCAGTCTGGC	5220
	TGTTCAGCAG	GCCCCATGCC	CCCCTTCCC	CAGTCTTGAG	GCCTGGGACT	AGGGCTGTCA	5280

	GGCACGTTTG CCACGTCTGC CCCTCTCTCC CCTGCGGCCA GCCCTCTACA GCCACAAGCC	5340
5	CGAGGTGGCC CAGTACACCC ACACGGGCCT GCTCCGCAG ACTATGCTCA TCACCGACAC	5400
	CACCAACCTG AGCGCCCTGG CCAGCCTCAC GCCCACCAAG CAGGTAAGGT CCAGGCCTGC	5460
10	TGGCCCTCCC TCGGCCTGTG ACAGAGCCCC TCACCCCCAC ATCCCCCGGG CTCAGGAGGC	5520
	TGCTCTGCTC CCCCAGGTCT TCACCTCAGA CACTGAGGCC TCCAGTGAGT CGGGGCTTCA	5580
	CACGCCGGCA TCTCAGGCCA CCACCCTCCA CGTCCCCAGC CAGGACCCTG CGGGCATCCA	5640
15	GCACCTGCAG CGGGCCCACC GGCTCAGCGC CAGCCCCACA GGTGAGAGGC CCTGGCTCCA	5700
	CCCCCTCCCT TACTGTCCCT GCCCCCTTCC ATGTTGGTCC CACCCCTTCT GTTGCTGTCC	5760
	GTCACTGTGG GGCTGTGCAT GCAGCAGGCC TAGGGCTGCT GTGAGGAAGC ACTGGCAGGC	5820
20	GTGGAAGGGT GGGGTGGCTT CCATGAATCC AGTGTTCACA GTAAGATGTA CTCAGGCCAG	5880
	TCCATGGGCG GCCGTGGACC CTGGCTGGGA GGCTCCCTTT GTTAAGAACC GAGGGTAGAG	5940
25	GTGTGACTTT GGGGTTCCCTG TTATGTGCTG TGATCCAGGA GGTGTGGCCC TGCCTCCCCA	6000
	TCCTGAGTAC CCCTAGGGAC AGGCAGGTGG GGTGGGTGTG GGTGCCTGGT GGGTGGCTAG	6060
	CAGCCTTGTGTT TGCCTCTGCA GTGTCCCTCCA GCAGCCTGGT GCTGTACCAG AGCTCAGACT	6120
30	CCAGCAATGG CCAGAGCCAC CTGCTGCCAT CCAACCACAG CGTCATCGAG ACCTTCATCT	6180
	CCACCCAGAT GGCCTCTTCC TCCCAGTAAC CACGGCACCT GGGCCCTGGG GCCTGTACTG	6240
35	CCTGCTTGGG GGGT	6254

(2) INFORMATION FOR SEQ ID NO:127:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 631 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:127:

	Met Val Ser Lys Leu Ser Gln Leu Gln Thr Glu Leu Leu Ala Ala Leu	
	1 5 10 15	
50	Leu Glu Ser Gly Leu Ser Lys Glu Ala Leu Ile Gln Ala Leu Gly Glu	
	20 25 30	
55	Pro Gly Pro Tyr Leu Leu Ala Gly Glu Gly Pro Leu Asp Lys Gly Glu	
	35 40 45	
	Ser Cys Gly Gly Arg Gly Glu Leu Ala Glu Leu Pro Asn Gly Leu	
	50 55 60	
60	Gly Glu Thr Arg Gly Ser Glu Asp Glu Thr Asp Asp Asp Gly Glu Asp	
	65 70 75 80	

Phe Thr Pro Pro Ile Leu Lys Glu Leu Glu Asn Leu Ser Pro Glu Glu
 85 90 95
 Ala Ala His Gln Lys Ala Val Val Glu Thr Leu Leu Gln Glu Asp Pro
 100 105 110
 Trp Arg Val Ala Lys Met Val Lys Ser Tyr Leu Gln Gln His Asn Ile
 115 120 125
 Pro Gln Arg Glu Val Val Asp Thr Thr Gly Leu Asn Gln Ser His Leu
 130 135 140
 Ser Gln His Leu Asn Lys Gly Thr Pro Met Lys Thr Gln Lys Arg Ala
 145 150 155 160
 Ala Leu Tyr Thr Trp Tyr Val Arg Lys Gln Arg Glu Val Ala Gln Gln
 165 170 175
 Phe Thr His Ala Gly Gln Gly Gly Leu Ile Glu Glu Pro Thr Gly Asp
 180 185 190
 Glu Leu Pro Thr Lys Lys Gly Arg Arg Asn Arg Phe Lys Trp Gly Pro
 195 200 205
 Ala Ser Gln Gln Ile Leu Phe Gln Ala Tyr Glu Arg Gln Lys Asn Pro
 210 215 220
 Ser Lys Glu Glu Arg Glu Thr Leu Val Glu Glu Cys Asn Arg Ala Glu
 225 230 235 240
 Cys Ile Gln Arg Gly Val Ser Pro Ser Gln Ala Gln Gly Leu Gly Ser
 245 250 255
 Asn Leu Val Thr Glu Val Arg Val Tyr Asn Trp Phe Ala Asn Arg Arg
 260 265 270
 Lys Glu Glu Ala Phe Arg His Lys Leu Ala Met Asp Thr Tyr Ser Gly
 275 280 285
 Pro Pro Pro Gly Pro Gly Pro Ala Leu Pro Ala His Ser Ser
 290 295 300
 Pro Gly Leu Pro Pro Pro Ala Leu Ser Pro Ser Lys Val His Gly Val
 305 310 315 320
 Arg Tyr Gly Gln Pro Ala Thr Ser Glu Thr Ala Glu Val Pro Ser Ser
 325 330 335
 Ser Gly Gly Pro Leu Val Thr Val Ser Thr Pro Leu His Gln Val Ser
 340 345 350
 Pro Thr Gly Leu Glu Pro Ser His Ser Leu Leu Ser Thr Glu Ala Lys
 355 360 365
 Leu Val Ser Ala Ala Gly Gly Pro Leu Pro Pro Val Ser Thr Leu Thr
 370 375 380
 Ala Leu His Ser Leu Glu Gln Thr Ser Pro Gly Leu Asn Gln Gln Pro
 385 390 395 400
 Gln Asn Leu Ile Met Ala Ser Leu Pro Gly Val Met Thr Ile Gly Pro

	405	410	415
5	Gly Glu Pro Ala Ser Leu Gly Pro Thr Phe Thr Asn Thr Gly Ala Ser 420 425 430		
10	Thr Leu Val Ile Gly Leu Ala Ser Thr Gln Ala Gln Ser Val Pro Val 435 440 445		
15	Ile Asn Ser Met Gly Ser Ser Leu Thr Thr Leu Gln Pro Val Gln Phe 450 455 460		
20	Ser Gln Pro Leu His Pro Ser Tyr Gln Gln Pro Leu Met Pro Pro Val 465 470 475 480		
25	Gln Ser His Val Thr Gln Asn Pro Phe Met Ala Thr Met Ala Gln Leu 485 490 495		
30	Gln Ser Pro His Ala Leu Tyr Ser His Lys Pro Glu Val Ala Gln Tyr 500 505 510		
35	Thr His Thr Gly Leu Leu Pro Gln Thr Met Leu Ile Thr Asp Thr Thr 515 520 525		
40	Asn Leu Ser Ala Leu Ala Ser Leu Thr Pro Thr Lys Gln Val Phe Thr 530 535 540		
45	Ser Asp Thr Glu Ala Ser Ser Glu Ser Gly Leu His Thr Pro Ala Ser 545 550 555 560		
50	Gln Ala Thr Thr Leu His Val Pro Ser Gln Asp Pro Ala Gly Ile Gln 565 570 575		
55	His Leu Gln Pro Ala His Arg Leu Ser Ala Ser Pro Thr Val Ser Ser 580 585 590		
60	Ser Ser Leu Val Leu Tyr Gln Ser Ser Asp Ser Ser Asn Gly Gln Ser 595 600 605		
	His Leu Leu Pro Ser Asn His Ser Val Ile Glu Thr Phe Ile Ser Thr 610 615 620		
	Gln Met Ala Ser Ser Ser Gln 625 630		
(2) INFORMATION FOR SEQ ID NO:128:			
	(i) SEQUENCE CHARACTERISTICS:		
	(A) LENGTH: 6433 base pairs		
	(B) TYPE: nucleic acid		
	(C) STRANDEDNESS: single		
	(D) TOPOLOGY: linear		
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:128:		
	CATGAACCCCC GAAGAGTAGT GTCTTCTCTC TGGACTAAAG CGGAACTGAG AACCGGTGGA	60	
	AAAGCCCCGC GCCTAGGCTG CAAGGCAGTG GCTTAACAAG TCCAAAGGTT AGGTGAAGTT	120	
	TGGCTGATAA GCAGAACCAAG TAAAAGAAGG TCTCTAGCCC CCCAGCGTGA GTACAATGGA	180	

	CCCTGGCAA	GCCCCGCTCC	CGGCCAGGT	CTTCTGCTCT	CCAGGTCTGC	CCCTCCGGCT	240
	CTCCCTCTCT	CCGGGTTTCC	CCCTCCCCAC	CATCATTGCA	ATCCAGCCGA	AAGCTGGGCC	300
5	CTTCCCACTA	ATTTGCATAT	CTTATATGGC	CTAATGGTGG	CGATCATGGC	AAGTTAGAAG	360
	TTTTCTGACT	CCTTCGGAG	GAGCCTCCGG	GACCCGGGG	AGTAACAGGT	GTCTGGAGGC	420
10	TGAAGGGTGG	AGGGGTTCCCT	GGATTGGGG	TTTGCTTGTG	AAACTCCCCT	CCACCCCTCCT	480
	CTCTCGCACC	CACCCACCCC	CTCACCCCC	TCTTTTCCG	TCCTTGAAA	ATGGTGTCCA	540
	AGCTCACGTC	GCTCCAGCAA	GAACTCCTGA	GCGCCCTGCT	GAGCTCCGGG	GTCACCAAGG	600
15	AGGTGCTGGT	TCAGGCCTTG	GAGGAGTTGC	TGCCATCCCC	GAACTTCGGG	GTGAAGCTGG	660
	AGACGCTGCC	CCTGTCCCCT	GGCAGCGGGG	CCGAGCCGA	CACCAAGCCG	GTCTTCCATA	720
20	CTCTCACCAA	CGGCCACGCC	AAGGGCCGCT	TGTCCGGCGA	CGAGGGCTCC	GAGGACGGCG	780
	ACGACTATGA	CACACCTCCC	ATCCTCAAGG	AGCTGCAGGC	GCTAACACACC	GAGGAGGCAG	840
	CGGAGCAGCG	GGCGGAGGTG	GACCGGATGC	TCAGGTAGGC	GCAGAGCCAG	GTGGAGGGGA	900
25	CCCACCCGAA	CCCCTGGAGC	CCCGGCCCCG	GGCCTGAGTG	ACACTGCGCC	CGACCACACT	960
	CGCCAAGCCC	GTTCCCCACC	AAAAAATTCC	CCCGGGGGGC	GCTCTGCTTC	TCTCCAACA	1020
	CCCGGACCC	TCCCAATCCC	TTAGCGGGAC	AACCCCTGCGG	CCCACCGGGC	TTCTTCTCCC	1080
30	CAGGCCAGG	CCATCGTCCT	CAGAAGAAAG	GGATGAGGTG	TACCGTACAG	GGGCAGTCAC	1140
	CTTCTCCTCT	GTTCAGCTTC	CATTTTGGCC	TCATGTCTAC	CCCAAAGTTG	TAGCTTAGAT	1200
35	GGGGGGAAAA	TTCAGAATTTC	TGCATAGACC	ATAGGTAGCA	CCCCCTAGAA	AAAGAATGTT	1260
	TCTCCCCAGA	TGTCTCCCAC	TAGTACCCCTA	ACCATCTGCT	TGTCTGTCTA	GTGAGGACCC	1320
40	TTGGAGGGCT	GCTAAAATGA	TCAAGGGTTA	CATGCAGCAA	CACAACATCC	CCCAGAGGGA	1380
	GGTGGTCGAT	GTCACCGGCC	TGAACCAGTC	GCACCTCTCC	CAGCATCTCA	ACAAGGGCAC	1440
	CCCTATGAAG	ACCCAGAACG	GTGCCGCTCT	GTACACCTGG	TACGTCAAGAA	AGCAACGAGA	1500
45	GATCCTCCGA	CGTAAGTGT	TTCATCCTGC	CTCTGCCCTCA	ACCTGAAGTG	ACCTTGCCC	1560
	TCTCACCCCA	TTGGCTGCCT	CAGTTCCCT	TTCATCGACA	AGGCCTTGTG	AGCACTTGGC	1620
	AGATATGAGG	AAGGTGGCAA	GTAGATTG	CCTTGGTGGT	TGCTGTACAA	TGGATTGGCT	1680
50	TCTGTATGT	TCTTCAGTCA	CAGCCCCCTT	GCTACCCAGC	CAGTTGCTCT	GAGGAGCCTG	1740
	TCAGTGTGAT	TGAGCTCACC	CACTTGACAT	CAAATACAGG	AGTTCAGGAT	GCAGAGTGT	1800
55	GCTTCATCTC	TGAAGGCCAG	TGAGCCAAAG	GGGAAAAAAT	AATAATTTC	TTAAAACAT	1860
	AGCTGGCTAT	GTTCAGCTC	CTTCAAAGAA	AGGAAAAGGG	TGGCTTGCT	GGAGCAACTG	1920
60	AGGTGGCAG	TAAGGGCCTG	TGCTGAGGGC	TCCCCATCTC	CAGCTCCACA	TGCAGTGAGA	1980
	GAAGGTTGCA	AAGCTTAGTT	AGACGAGGGG	AATAAACCTG	TCTTCGTCCG	TTGTCTGTCT	2040

	GTCTGTCTGT	CTGTCTGCTG	AGTGAAGGCT	ACAGACCCTA	TCAAATCTAC	TCCTTTCTCT	2100
5	TTTCAGAATT	CAACCAGACA	GTCCAGAGTT	CTGGAAATAT	GACAGACAAA	AGCAGTCAGG	2160
	ATCAGCTGCT	GTTCCTCTTT	CCAGAGTTCA	GTCAACAGAG	CCATGGGCCT	GGGCAGTCCG	2220
	ATGATGCCCTG	CTCTGAGCCC	ACCAACAAGA	AGATGCGCCG	CAACCGGFTC	AAATGGGGC	2280
10	CCCGGTCCCC	GCAAATCTTG	TACCAGGCCT	ACGATCGGCA	AAAGAACCCC	AGCAAGGAAG	2340
	AGAGAGAGGC	CTTAGTGGAG	GAATGCAACA	GGTAACACCA	CCAGAAGCTC	AGGTGGGCAG	2400
15	GTGGGCAAGT	ACACAGACCC	AGGAACCCCTC	CCCTCGGTCC	TGGGATATTG	AGACACTAGT	2460
	TATAACAGATA	AGTGTGGCTA	AATCAGAGCT	TCTCAAAGTA	TGTTCCACAG	TGATTGTGTG	2520
	TTTTGGGCCA	AGCACCAACA	AGTCCCCCG	CCCCCCTTCA	CTCACCATCT	CCCCTCCATC	2580
20	CATTCCCAGG	GCAGAATGTT	TGCAGCGAGG	GGTGTCCCCC	TCCAAAGCCC	ACGGCCTGGG	2640
	CTCCAACTTG	GTCACTGAGG	TCCGTGTCTA	CAAATGGTTT	GCAAACCGCA	GGAAGGAGGA	2700
25	GGCATTCCGG	CAAAAGCTGG	CCATGGACGC	CTATAGCTCC	AACCAGACTC	ACAGCCTGAA	2760
	CCCTCTGCTC	TCCCACGGCT	CCCCCCACCA	CCAGCCCAGC	TCCTCTCCTC	CAAACAAGCT	2820
	GTCAGGTAAG	CAAAGGTTGG	GCCTCACTGC	CTCGGCAACC	CAACCATCCT	GGTTCTTGCC	2880
30	ACGGATCTTA	TCTGGTTAA	GGGTTTCAG	AGGAGCAAAC	GCTTTGAGA	TGATCCTAGG	2940
	GCCGCTCTCT	CATTGCCAGA	ATATACTCCC	CTGGAAATAA	TGTGTGGCTC	TGATCAGTTC	3000
35	CAAGGCAC TG	GGGATACATC	AGTGAACAAA	ACAAACGAGA	TAAAAATTTC	CTGCCCTCGT	3060
	GGCGCTTACA	TTCTAGAATT	AAATAGAGAA	CATGCCATAT	TTACCCCTGGA	GAAAGCAGC	3120
	CGATATTTCT	TGTGGGTGGA	CAGGGGAGGA	GAAAGCAACT	TTATTTCTT	ATTACCCACC	3180
40	CTTGAAAACA	AGAGGTGCCG	AGTCATTGTT	CCAGGACCCT	GGTGGCACTA	ATGTTCCCTA	3240
	CTGGGTTTGT	GTTGTTTGC	AGGAGTGCAG	TACAGCCAGC	AGGGAAACAA	TGAGATCACT	3300
45	TCCTCCTCAA	CAATCAGTCA	CCATGGCAAC	AGGCCCATGG	TGACCAGCCA	GTCGGTTTTA	3360
	CAGCAAGTCT	CCCCAGCCAG	CCTGGACCCA	GGCCACAATC	TCCTCTCACC	TGATGGTAAA	3420
	ATGGTGAGTA	CACCTGGGCC	ATTGTCGCTC	TGGAGCTGAT	AAGATAAGAG	GCAAAACAAA	3480
50	CACAAC TTCT	CACAAGGCCT	GCCTCAAACA	ATGAACCATT	GTAGCCCCAT	AGGGAAAAT	3540
	GAGGGCTGTC	CAGAGTCGGA	AAGGAGAGGT	AGTGCTGGTG	ACCCACCCCTT	TGGGGGTAG	3600
55	AAAACCCAAA	GTGATGGAT	TACAGGGGTG	AAGCACCATG	CCCAGCCAAT	AATTGTTATT	3660
	GAGTGAATGA	AGGAATGAAT	TTGAGAACTA	GTCATGCCAA	GGAATCGCTA	AGTCACATCG	3720
	TGTTGGAAAC	TGCTCTTGT	GGTCCAAGTC	CACCCATGTT	TCTCTTGT	TTTTCTCTCC	3780
60	ATCAGATCTC	AGTCTCAGGA	GGAGGTTGC	CCCCAGTCAG	CACCTTGACG	AATATCCACA	3840

	GCCTCTCCC	CCATAATCCC	CAGCAATCTC	AAAACCTCAT	CATGACACCC	CTCTCTGGAG	3900
	TCATGGCAAT	TGCACAAAGT	AAGTTCTATT	CTTGGTTGGA	AAACCTGGGG	GCAGGGAGAA	3960
5	GAAGAATGGG	AAGCAAATTA	ATGTGGTGAA	AAATAACTGT	AGGTCTCCTT	CAAACTCACC	4020
	CACAACTAGT	AAATTGGTT	TAACTTCTTT	AGTTTCTCAT	CTGTCTCCTT	AAATCCAATA	4080
10	TTTGGATTGT	TTAGCCTAAA	ACAAGAAAAA	ATTGTGGAAT	GGATTGGAAT	CCTGGTCACA	4140
	GTTTAGCAGC	TGTGCATCCT	GGGTCAAATC	ATTGAACCTA	TGACTCTGGG	AGACTCTCAG	4200
	GCTTTAACATCA	GATCTGTTA	ATGCCCATCT	CCAACCCACA	ACTCATTGTG	GAACTTGAGC	4260
15	AAGTAAATTA	ATATCTCCAA	GTCTCCGTTT	CTTTACACTT	GCCTCCCATG	GAATCTCCTA	4320
	TGTAACAGGC	TCAGCCCCGT	GAUTGGGACA	TTGAGCGGGG	GCTCAAATGA	TGGCATCCAT	4380
20	CCACCTCTCC	TTATCCCAGG	AGCTGTCTGT	GTCTTTCCCT	CTTGCTCCCA	CAGGCCTCAA	4440
	CACCTCCCAA	GCACAGAGTG	TCCCTGTCAT	CAACAGTGTG	GCCGGCAGCC	TGGCAGCCCT	4500
25	GCAGCCCCGTC	CAGTTCTCCC	AGCAGCTGCA	CAGCCCTCAC	CAGCAGCCCC	TCATGCAGCA	4560
	GAGCCCAGGC	AGCCACATGG	CCCAGCAGCC	CTTCATGGCA	GCTGTGACTC	AGCTGCAGAA	4620
30	CTCACACAGT	AAGGACACGG	GCATGTGGAG	GGAGGGAGCA	CTCAGGACCC	TCAGTGGCCA	4680
	ACCACTTTCC	CTCTCTGGGT	CTGAACTTTC	TCGGAAGTTT	ATTGGCTTGG	TCACTTTCC	4740
35	CTGCCTATGA	TCAACCGACT	AAGACAATT	CTCAAGCATA	ACTCTTGAGT	GTTGCTGTAC	4800
	CTTTTCTAGT	CCTCTCTCT	ACCCCTGAGA	TTCCCAGGGA	AGGGTTTGAA	TGACCTTTGC	4860
40	TCCCGTTCCG	TACCGGAGGC	CTCCCTGGTA	GGAAATGTGT	TCTGAGAGCA	GGTGGTTCT	4920
	CCCTCACAGC	CAAGCATCCA	CATGCTTCG	GGAGTTGGTT	ATGTGACTTG	GAATTACAT	4980
45	GAATCTTATG	GATAACTAAT	ATGAGAAATC	CCCACTATAA	CCACCAGCCC	TTTTATCTAC	5040
	CTGAGGAGAT	GGGAGCTATG	GTGTGGGATG	GGGGCTCTGT	ACCTGTGTCT	TTGCCTGTGT	5100
50	ATGCACCTTG	ATTCTGTCTT	CACTCTGTCT	CTCCAGTGT	CGCACACAAAG	CAGGAACCCC	5160
	CCCAGTATTC	CCACACCTCC	CGGTTCCAT	CTGCAATGGT	GGTCACAGAT	ACCAGCAGCA	5220
55	TCAGTACACT	CACCAACATG	TCTTCAAGTA	AACAGGTAAT	GCCAGCAGGA	TATGCGGGGG	5280
	TTGGGGTGTG	GGCAGGGTGT	GATAAGGCCA	TGGATGTGCA	AAGGTTGTGG	CAAGCATGGA	5340
60	CTCGGCCAGA	AATTATATCC	TCTTGCTGG	TTGAGTTGGG	CATCATCTCC	CTTAGAGAAAG	5400
	CCAAACTAAT	GGCCCATGAC	CCTGCCAAAT	GACACAGCTG	AGCACCCCTCT	CTCCTCTCTC	5460
65	TCTGCAGTGT	CCTCTACAAG	CCTGGTGATG	CCCACACACC	ACTTACTTCG	TGCGCAACAA	5520
	CAAGGACCCCT	GTTCCTCACA	CCATCACCC	CTGGGCAGCT	GTCATGGAAA	AGCCCAGTGA	5580
70	CCTGACCAGC	ACCTGCGAGA	GGTCCCTGCT	ACCTGACGGA	CGTCCTGCTG	GCACCTCAGA	5640
	CAATCCACTC	TCAGGAGGCG	CAGCCCGAAG	CCCAGTTCC	CTTCTATGCA	GTATTGCCAC	5700

	AATGCCTCTC CCACGATGTC AAGGACTCCT GTCTGTCCCTG GAGGTGGGAG ACAAGGAACC	5760
5	ACCGAAGAGG AAGCAAGAAA GCCGTACTGT CTATGTTGTG ATCCTTCATC GAACAAACTG	5820
	ATGCGAAAAC TTGAATCTGT TACTGAAATG AGGAGAGAAG GACATGTGCT ATTGAACCTGA	5880
	GCCAAACACA CTGTAAATAT CCACAGACTC CCTCCCCCTGC CCCCATCCCA CATGATCTTG	5940
10	AGATTCTTT TAAAGAAGTA AATTGTCCA ATGGCTGTAA ACTATAAACT ACTGTAATTA	6000
	AGTGCAATTT CCCCTCTGTG TCCTCTCCCC TCTGCCCTGT ATATAAACT AAAGTGTCTA	6060
15	TTAGTTTCT TTGTAAAGGT CAGAGTCAAA ATTTCAAAAG TGATCTGTCC CCTCTCCCCT	6120
	CATGGAGAAA CATCCTAAGT GGGAAAGTGAA GCCCCCTGTG CTCTCCCGCG GGCCTGGACA	6180
	CTTATGGGGA CAGCATAACCT TGGACTGACT ACCAGCTAAC TCCAGTCTCC TGACATTAAG	6240
20	ACACACCTCT GGATCCCTGG AGGGGCTGAA TGTAGTGTGT CAGAGTAACA TGCCAGCTTC	6300
	CTGTGGGCCA GGAGCTCAGC CTGCACTCCC TAAGAAACCC CAGGGCAGGG AAACCTGGCTG	6360
25	TTTGATAGCA GAAGAAAAAG TTGCAGTCTC AAAAGCCTTC CATTAAAACA ATTTATTTA	6420
	TCACTAAAAA AAA	6433

(2) INFORMATION FOR SEQ ID NO:129:

- 30 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 609 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:129:

40	Met Val Ser Lys Leu Thr Ser Leu Gln Gln Glu Leu Leu Ser Ala Leu 1 5 10 15
	Leu Ser Ser Gly Val Thr Lys Glu Val Leu Val Gln Ala Leu Glu - 20 25 30
45	Leu Leu Pro Ser Pro Asn Phe Gly Val Lys Leu Glu Thr Leu Pro Leu 35 40 45
	Ser Pro Gly Ser Gly Ala Glu Pro Asp Thr Lys Pro Val Phe His Thr 50 55 60
50	Leu Thr Asn Gly His Ala Lys Gly Arg Leu Ser Gly Asp Glu Gly Ser 65 70 75 80
	Glu Asp Gly Asp Asp Tyr Asp Thr Pro Pro Ile Leu Lys Glu Leu Gln 85 90 95
55	Ala Leu Asn Thr Glu Glu Ala Ala Glu Gln Arg Ala Glu Val Asp Arg 100 105 110
60	Met Leu Ser Glu Asp Pro Trp Arg Ala Ala Lys Met Ile Lys Gly Tyr 115 120 125

Met Gln Gln His Asn Ile Pro Gln Arg Glu Val Val Asp Val Thr Gly
 130 135 140
 5 Leu Asn Gln Ser His Leu Ser Gln His Leu Asn Lys Gly Thr Pro Met
 145 150 155 160
 Lys Thr Gln Lys Arg Ala Ala Leu Tyr Thr Trp Tyr Val Arg Lys Gln
 165 170 175
 10 Arg Glu Ile Leu Arg Gln Phe Asn Gln Thr Val Gln Ser Ser Gly Asn
 180 185 190
 15 Met Thr Asp Lys Ser Ser Gln Asp Gln Leu Leu Phe Leu Phe Pro Glu
 195 200 205
 Phe Ser Gln Gln Ser His Gly Pro Gly Gln Ser Asp Asp Ala Cys Ser
 210 215 220
 20 Glu Pro Thr Asn Lys Lys Met Arg Arg Asn Arg Phe Lys Trp Gly Pro
 225 230 235 240
 Ala Ser Gln Gln Ile Leu Tyr Gln Ala Tyr Asp Arg Gln Lys Asn Pro
 245 250 255
 25 Ser Lys Glu Glu Arg Glu Ala Leu Val Glu Glu Cys Asn Arg Ala Glu
 260 265 270
 Cys Leu Gln Arg Gly Val Ser Pro Ser Lys Ala His Gly Leu Gly Ser
 275 280 285
 30 Asn Leu Val Thr Glu Val Arg Val Tyr Asn Trp Phe Ala Asn Arg Arg
 290 295 300
 35 Lys Glu Glu Ala Phe Arg Gln Lys Leu Ala Met Asp Ala Tyr Ser Ser
 305 310 315 320
 Asn Gln Thr His Ser Leu Asn Pro Leu Leu Ser His Gly Ser Pro His
 325 330 335
 40 His Gln Pro Ser Ser Ser Pro Pro Asn Lys Leu Ser Gly Gly Lys Gln
 340 345 350
 Arg Leu Gly Leu Thr Ala Ser Ala Thr Gln Pro Ser Trp Phe Leu Pro
 355 360 365
 45 Arg Ile Leu Ser Gly Leu Arg Val Phe Arg Gly Ala Asn Ala Phe Glu
 370 375 380
 50 Met Ile Leu Gly Pro Leu Ser His Cys Gln Asn Ile Leu Pro Trp Lys
 385 390 395 400
 Gly Val Arg Tyr Ser Gln Gln Gly Asn Asn Glu Ile Thr Ser Ser Ser
 405 410 415
 55 Thr Ile Ser His His Gly Asn Ser Ala Met Val Thr Ser Gln Ser Val
 420 425 430
 60 Leu Gln Gln Val Ser Pro Ala Ser Leu Asp Pro Gly His Asn Leu Leu
 435 440 445

Ser Pro Asp Gly Lys Met Ile Ser Val Ser Gly Gly Gly Leu Pro Pro
 450 455 460
 Val Ser Thr Leu Thr Asn Ile His Ser Leu Ser His His Asn Pro Gln
 465 470 475 480
 Gln Ser Gln Asn Leu Ile Met Thr Pro Leu Ser Gly Val Met Ala Ile
 485 490 495
 Ala Gln Ser Leu Asn Thr Ser Gln Ala Gln Ser Val Pro Val Ile Asn
 500 505 510
 Ser Val Ala Gly Ser Leu Ala Ala Leu Gln Pro Val Gln Phe Ser Gln
 515 520 525
 Gln Leu His Ser Pro His Gln Gln Pro Leu Met Gln Gln Ser Pro Gly
 530 535 540
 Ser His Met Ala Gln Gln Pro Phe Met Ala Ala Val Thr Gln Leu Gln
 545 550 555 560
 Asn Ser His Met Tyr Ala His Lys Gln Glu Pro Pro Gln Tyr Ser His
 565 570 575
 Thr Ser Arg Phe Pro Ser Ala Met Val Val Thr Asp Thr Ser Ser Ile
 580 585 590
 Ser Thr Leu Thr Asn Met Ser Ser Lys Gln Cys Pro Leu Gln Ala
 595 600 605
 Trp

(2) INFORMATION FOR SEQ ID NO:130:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 10014 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:130:

45	TGGGTTGCCT GTGACTGCAC TGGCGATACC CCCACAAAGC CCACTCTGAA GGTAGGAGAC	60
	GGGTGGAGAG AAACAGGGGG ATGGCAAGGG GGATACGAAA CAGGGAGAGG GAGGAGGGGG	120
50	AAGAGGATGG ACGTCTACCA GGCCCCACTT GGTGCTTGAT TTATGCCATC TCATTTCCCTT	180
	CTCAAACAC CCTTGAAAGT TGATTGTACA TTTTACAGAA AAGGAAACTG AGGCTCGGAG	240
	AGGAGAACATCA TTTACCCAAG GTCCCAGTTA GTAGACGGTA GGTGCCTGAA TGTAAATCCA	300
55	GGTCTCTGCC TGCTCCGGGA GGGGGTGGGG GTGAGGGAAA CAGGAGAACATG TGATGGGAAA	360
	ATCCGAGATG GAGCCAGCCT GGGCCAGAAA CACTGGGAGC TGTGGGAGAC GGAGAGGGGC	420
60	AGGGTGGGAT CACAGGGAGC AGGAGCGGGG AATTGGAGGT GAATCTGGCC CTCCCAAAC	480
	TCCAGTCCAT TCTGCTCCCA GGGGAACCGG GAAACTGCGG GGGAACTGGA AGGGAGCTCC	540

	CAGAACAAAGG ATCCAGAAGA TTGGCATCTG GGGCCTGGGA TTTAGGTTTC TAAATCGTGG	600
5	GCCATGGGGC AGCCTATCT CTGCAAAAGC ATTGAGGGTA GAAGTCAATG ATTTGGGAAG	660
	TTATTGAATT AGGGGATCTC GGAGGTAGGC TGTCAGTGCC TGATAGTATC AGTTAGAATG	720
	CCTGACTTGG GGTGACAATG GCTTGGAGGG GTGGGTGAGT CAAGGGTCAA ATGAGTGCCC	780
10	GTGAGTCATG ATGCCTGCCT TGTACAATTG ATAACGTAAAC ATCGGTGAGT TAGGGCCCCA	840
	GCAGTTGTAA TTAGCACCCC GGGTGTCAAG CAGAAACCAA CAAACAGCCA AATCCCTGCA	900
15	GCCCCGCCA GCCTATCCAC CGGCAGGGGA CCGATTAACC ATTAACCCCC ACCCCTCCCC	960
	GGCAGAGCCT CCACCCCTTC ACAGAGGCTA GGCCAAGACT CCCAGCAGAT CTTCCCAGAG	1020
	GACGGTTGAA AAGGAAGGCA GAGAGGGCAC TGGGAGGAGG CAGTGGGAGG GCGGAGGGCG	1080
20	GGGGCCTTCG GGGTGGCGC CCAGGGTAGG GCAGGTGGCC GCGGCGTGGA GGCAGGGAGA	1140
25	ATGCGACTCT CAAAAACCC CGTCGACATG GACATGGCCG ACTACAGTGC TGCACCTGGAC	1200
	CCAGCCTACA CCACCCCTGGA ATTGAGAAT GTGCAGGTGT TGACGATGGG CAATGGTAGG	1260
	TGGGGCAGA TGTGCCCAGG TGTGCCAGTG GGGCAGGTG TGCCCTGGTC CAGGAGCAGA	1320
30	TCTTTGGCAC TCAACTTTGG GGTGGGAGGA GAATGATACA AAATGGTAGG TTGGTCTAC	1380
	AGGCCAGCAC AGGTGTTGCC AAGTGAAGCC CATGTGCCA GGCACAGTGA TCACAGGCAT	1440
	TCTGGGTGAA GGGAGGCCTG CAAGGGCCAA TTTCCAGCAA AAGTCGATCC CGGCTATTCC	1500
35	TCCCCAGGCC TTCCAGTCCT CACTGCCTCA CAGTGGCTCT GCTTGGCGCT TGGCACAGTG	1560
	ACATGATGGT GAGCTCCCCC TTGGTGCCCA GCTCCAGCGA TTCAGCCCAG CACGGCCCCT	1620
	TCGTGAACCC CTTGGGCCTA GGTCAGAGA GACGGCAAGG GATGTTGTAT CCCTGGAGAT	1680
40	GGTGGTTGGA GACATAACCG CATTCTCGG TGTCTTGGA ACTTTCTCTAG GGAAATGAAA	1740
	TTGGCACTTA GGGAAAATGG AGCTCTCAGG GAAGTTTGCA TAACTACGAA GCCAACTCAG	1800
	CACTGTGTGT GTTGTGTGT CGTTCTGTG TGATAGTGTAG TTTCCATGTA GGTTGTATGG	1860
45	GTGGGGTGAT GCCTTCAGGA ACCCATTGTC ATATGTGTGT TCATTTGTCT CTGTGTGTGA	1920
	GTTCTGGTC TATTTCTCTT TGTATTCTT GAGTGGGTCT GTGTTGTGT CTTAGGAGTT	1980
50	GCCCGTGTG ATCTTGCTTA TGTATGTAAG TGTGTATGTG TGTGTACTTG TGTCTGTGGA	2040
	TGTTTGACCA TGTGTGCTGT GTGTGCGGGT CATAGAGCAC ATGCGTTGT GCATGCGGAC	2100
55	CTGTTGGAGT GCCCTGTTCT TCCTGCATCT TTATCCTGTA TGGCGTTTT GTCTGTGCCC	2160
	CATATTGTA CCTGCTGTGT ATATATGCAG TTCCCTGTGC TGCAGGGCGGG GGTCAGCGGT	2220
	CTCTGGTGTG CACGACTGCA CAGACCCAAA TGCAGGACTC TGTGTTGCC ACTCACCAAG	2280
60	TGAGATTCAATCAGCAACA TGTCCGTTTG TCTCTGAGCA GATTTGTTTG CCGCTGCGTC	2340

	TCGCCAGATT GAGGCATCCC CTCCGACATC ACTGGAGCAT ATCTGGAGGG GTGGACAGTT	2400
	CTCCACAGGG AGGTAGGGGA AAAGAGGAGG CCCGGAAACC CCTCCTGGAG GGAAGAGCCC	2460
5	CATCGGTCCC AGGCCAGCCT CAGAGGAGAG GGGGCAGGCA GCTGGCTGAG GTCAGCCTGC	2520
	CACCCCTGCTT CCTTCTGTGT CTTGGAGCCA CTCAGCCAGT ATGAGGCTGC AGCTCCAGCT	2580
10	GAGGTCTGGA ATCTTGTGGT CAGCTCAGCT AGGGTGAGGA GGCAGCTGCT GGGCACTGCT	2640
	TGTTGTCAGC TCAGCAGGTG CTCACCTGCC CCTGCCGTCC AGTCACGTGT GACCTTGGC	2700
	ATGTCACCTC CCCTATCCTG GCTTCTGTAT CTTCTACAAA ACAGGCTTCA TTCCCCCAGG	2760
15	CCTGCTGGCT GGACGGCTTT TAGGCCTGTC TGAGGACCAC GCCAGGAGCG CAAGGAAAAA	2820
	ACACACCAGA GATCCCCTTG CGAGTTAGGA GGCCGGCTCC CACCCCAGAA GGTGGCCAGG	2880
20	TTTTCATGCC TTCCTAGAGA AAGCTGGGC TGGTGGCCTC CACCACAGGG AGACGCAGAC	2940
	CCTCAGAAC AAGTCTGTGA AGTCACAACC AGCCCCAGTT TACAGATGTG AAACTGAAGC	3000
25	TCCAAAAAGT CAGGAGGTCA CTGAGTGGGG AGGTGATGGA GTGGGAACAG CCCCCAGATC	3060
	TGGCTGAGGC CGAAGCCCTG GAGAGATCCC CGCAAGGCTC CCTTAGATGC CTGACATTCT	3120
	GCTCTCCTG AAGCCTCACT CCCTCTCTC CTGGCGCAGA CACGTCCCCA TCAGAAGGCA	3180
30	CCAACCTCAA CGCGCCCAAC AGCCTGGGTG TCAGCGCCCT GTGTGCCATC TGCGGGGACC	3240
	GGGCCACGGG CAAACACTAC GGTGCCTCGA GCTGTGACGG CTGCAAGGGC TTCTTCCGGA	3300
	GGAGCGTGCG GAAGAACCAC ATGTACTCCT GCAGGTGAGG AGCCTCAATT TCTTCAGCTG	3360
35	GGAAATGGGC ACACTTGGC TCATGGCCCC AAGGTCTGTC TTCTCCCTGA GTGGGTAGGT	3420
	CCCAGAGACA GCTGCCCTTC AGGGCCTTCA AGGCTCTTCT GGTTTGTAA AAGACTTTGT	3480
	GAATCCAAGA AGAGCATCTA TTCTAGGAAC CACATTACT GATCATCAAG CTACTGGCTG	3540
40	CCGTTTATTG AGCTCTTATC ATATGCCAGG CACAATACTA AGTCTTTGTG TGTATTTACG	3600
	TACTCCAGAG GTCAAGGTTTC CCAACTCAGC TCTAACACCA ACCAGCAGAG CGACCCAGGA	3660
45	CCACATGTTG CCTCTCTGAG CCTCAGTTT CCCATGTTA GCAGGACAGG ACTGGGCTCT	3720
	TAGAGAGTTC ATAGCACCTT TCCAGCTCCT GGTGGGTTCA AGAGAGAACT CCCGGGATGA	3780
	AGAGATGAGA GCACTGAGGT TGGGGGTCA ACTGGATAGC CAGGGCCCTA GTTCTGTCCT	3840
50	AAGAGGAGGA AGTTGTGTCT TCTCCATCCA ACCATCCAAA GCCCTCCCCA GATTTAGCCG	3900
	GCAGTGCCTG GTGGACAAAG ACAAGAGGAA CCAGTGCCGC TACTGCAGGC TCAAGAAATG	3960
55	CTTCCGGGCT GGCATGAAGA AGGAAGGTGA GCCTCGGCC CCCCCGCC ACCACCACTG	4020
	CCCCACCTGC ACCCACAGCT CCCCCACAGT CATTTACAAC TGTAGCCACA CTTTATGACT	4080
60	CAGTGGCAGG CCCCAGGGTG ACTGGCTAAT GGCTGAGAAG AGGGAGGGCC TGGAAATCTG	4140
	ACCATAGGGA GCGGCTGGGC TTGGTCTTGA GAAAGATTCT CCCACTCCTC ATCAGTCACA	4200

	GACACCCCCA	CCCCCTACTC	CATCCCTGTT	CTCCCTCCTC	ACCTCTCTGT	GCCTCCTCAC	4260
5	CCGTCCAGAA	TGAGCGGGAC	CGGATCAGCA	CTCGAAGGTC	AAGCTATGAG	GACAGCAGCC	4320
	TGCCCTCCAT	CAATGCGCTC	CTGCAGGCGG	AGGTCCCTGTC	CCGACAGGTA	CCGGGGTGAT	4380
	CCTGCCACCC	ACCCAGGGAT	CCCCCACACT	ACAGAGGAGC	TCACCTCCTC	CACCTCCATT	4440
10	CTCCCCAGCC	AGGCCCTGGA	GCAGCTGACC	GGAGGGGCCT	CAGATATTAC	AGAAGGGACA	4500
	CTGAGTGCAG	TTTCACATGG	CCCAGTTTGC	AGCAAGGGCA	GGAATCGAAC	CTGGCGCCCT	4560
15	GGGGCACTTT	CTAATTCAATC	CTACTGCCTG	CATCCCACAG	GCCAAGCAGA	GTCTTCACCT	4620
	TCACTGAGGG	CCTGCGATCA	GCTCAGCTCC	GAGAGAACAG	AGCAGTGGCT	CAGTGGAGAG	4680
	AGGTGGCAAA	GTGGGGCCA	GCCCTTCCCT	TGCTGAGTGA	CCTTGGGCAA	GTCACAGCAC	4740
20	CTCTCTGAGC	CATGGTGCC	TCATTGTCAG	AAAAGGATGA	TGATTTTTTG	CCCTGCTTCT	4800
25	CCTCTAAGGC	TGACAGACTC	CTTGGGGCTC	TAAAGCTGTT	CTCCCTCATC	CCTGCCTCCT	4860
	CCCTCCCTCC	GTAAAAACCC	TGAGCTTCCT	TCAGAGCTGG	AGGGCACCCA	CTATCCAGCC	4920
	CCCTCCCCAC	ATCTGATTCC	AGGGAGGGGG	CTCTGTGCAG	GGGACAGAGA	ATGCAGGAGG	4980
	GCCCGGACAT	CTCCAGCATT	TTCTTCCCTG	TATCTCTCGA	AGATCACCTC	CCCCGTCTCC	5040
30	GGGATCAACG	GCGACATTG	GGCGAAGAAG	ATTGCCAGCA	TCGCAGATGT	GTGTGAGTCC	5100
	ATGAAGGAGC	AGCTGCTGGT	TCTCGTTGAG	TGGGCCAAGT	ACATCCCAGC	TTTCTGCGAG	5160
	CTCCCCCTGG	ACGACCAGGT	GAGGATGGGC	GTGGATGGTG	GGCAGTAGTG	GGCAGTGGC	5220
35	GGGGCAGCCA	GGGGGCTGCT	GGCCCACCTG	GGATATAAGCC	GTGGACTGGC	TTGATTTTAT	5280
	TTTATTTAAC	AAAATATGTA	GTGCACACAC	GTGTCTGAAA	CTTTAAATCA	CCTTACAAAT	5340
40	ATTAACACTAG	TTAGCTCCTC	CAACAACTCT	ATGAGGTAGG	TACTAAGGTA	CTATTATTAC	5400
	TGCCATCTCA	TAGGTGAGGA	GATTGGGC	CAGAGAGGTT	AAGTAACCTG	CTCAAGGTCA	5460
	CATAGCTACT	ATCCAGCATA	GCTGGGATT	TTACAAAGCA	CCCTTCATAA	TTCTCCATAG	5520
45	CTGGTCCATG	GGTGGGAATT	TGGGACCCAC	AGTTTTGGAA	CTTTTTGGGA	TCATAGACCT	5580
	TTTGAGAAT	CTCAAAAAAG	AAAAAAAAAG	CACACAGAAT	GTGCTTACA	GTTCATCAG	5640
50	GCACACAGAA	GAGGCCAGC	ACGAAGCAGT	TTCTTGCCCA	AGGACACAGC	AGTTCAAGGA	5700
	CAGAGTCAGC	GCGAGGTCTC	TCAGCTCTGA	GCACATGTT	TTTCCCCCTTC	CAGGTTCTA	5760
55	GTGGGGGGGG	TAGTAGTTTT	ATGATGCCA	TTTCACAGTT	CAGGCAGGTA	GAGGCAGAGG	5820
	GGAGCATTAA	GCTGACTTGC	CCAGCGTCAC	TGAGTTGGCT	ACGGGCAGCC	TTCCCAAGGG	5880
	TACAGATGGC	AAACACTGTT	CCTTCTCTCT	TTCAGGTGGC	CCTGCTCAGA	GCCCATGCTG	5940
60	GCGAGCACCT	GCTGCTCGGA	GCCACCAAGA	GATCCATGGT	GTCAAGGAC	GTGCTGCTCC	6000

	TAGGTGAGGC GGCTGCCTGC CCTGGCCAGG GCTCCAGGGA GGGTATGCCT AGCATGGCAC	6060
	TCACCCAGGC AAGGAGATT ACATGGTGGC ATGCAAGGGT GAGGGAGACT AGTCAGGAGT	6120
5	GGCCCTGTCC TCAGGCTTGC ATTGGAGGGC TCCAGGACTC AGTTTCAAC TGGGTACCCC	6180
	ACTCAGATGC AAGGAAATGT GGATGCAAGT CACCAAATTC CCAGCATTGA AGTCAGAGCA	6240
10	CGATCAGGGT TATCCCTGGA ATTACCTGTG CATCCTTTT TCTTTGACA GAGTCTTGCT	6300
	CTGTCACTCA GGCTGGAGTG CAATGATGTG AGCAAACACT ACCTATTTA ATATAACAAT	6360
	GCTATGAGGG AGCTCGATTA TTTATCCTCA TCTTATAGAT AAGAAAATG AGGCACAGAG	6420
15	AGGTTAACGTA ACTTATCCAA CTATAACCAG CTATCAGGGG CAGAGCCATT TAAGCAGGGC	6480
	AGTGCAGTTC CAGAACCTGG TCCTTTAACCC TTGATGCTTT GGTGCCTATC AGGTGACCTT	6540
20	TGAATGTCAT CGATCTTGTG AGTCATGTTG GTAAATGGAG CTTGGTCAT GTGAAAGAGG	6600
	TCCTAGAAAG CCAAGTTCCA AGCTCAGCCG GATGACTCAA GGCAGCTTAT CTTCTGAATC	6660
25	TGGGCCTCAG CTTCCCTTACC TGTGAAATGG GAGTCACCAT CCCTGCAGGT CCTCCTCCCA	6720
	CAGGCACCAAG CTATCTGCC AACTTAAAAG CCAAAACTAG AGGAGAGGGG TCAACCCAAG	6780
	GTGACTTCCC ATCCTCCCTC CCTCCCAACC CTTCCAGGCA ATGACTACAT TGTCCCTCGG	6840
30	CACTGCCCGG AGCTGGCGGA GATGAGCCGG GTGTCCATAC GCATCCTTGA CGAGCTGGTG	6900
	CTGCCCTTCC AGGAGCTGCA GATCGATGAC AATGAGTATG CCTACCTCAA AGCCATCATC	6960
	TTCTTGACC CAGGTACAGT GCACACCTCC TAAGCCATCC CTGACTCTCT CTCCAGAACG	7020
35	CTCTGCCAGA CTTCTCCTAT TGGGTTCTGT ACACTGAGTT CACAGCCTCA TCTCATGTTA	7080
	ACGACAGCCA GGAGAGGCCG TTTTCATTAA ACAGATGAGG CAAGTCAAGA TTTGAAGAGA	7140
	CAATATGGCC GGGCGCAGTG GCTCACACCT GTAATCCCAT CACTTGGGA GGCTGAGGCG	7200
40	GGCGGATCAC CTGAGGTCAAG GGGTCAAGAT GAGCCTGGCT AACATGGAGA AACCCCATCT	7260
	CTACTTAAAA GTGGCTCTGC CAACAACTGG CTGTGCGACC CAGGACAAGT CCTATCTTG	7320
45	CACTGTGTCT GGGTTCCCC GTGTGTAAGA TGAGGCGGTT GCTAGGTGCT TATTGGATGC	7380
	ATTCCCTCAAG TCCCGCCCTC CATCTCCTAT TCCCCTCTCT TCTGGTTTAG TGCTTTAGGA	7440
	AATGTGGCAG AAATCTTTT CTGCCTGTGT CTAGGAAATC ATAATTGATG CTGGCGTACC	7500
50	CTGGTTGTTG AGGTCCCTGA ATCCTTGTGC CCACACTGCT GAAGACTCCT TGTGTGACAC	7560
	AAGTCAGGGG ACATCTGGGT CTTGACTCCC CAGATGCTCC AGCTGGACCC TGCTGCCCTC	7620
55	CCTGCCAC CCTCTTCCAT TGTAGATGCC AAGGGGCTGA GCGATCCAGG GAAGATCAAG	7680
	CGGCTGCGTT CCCAGGTGCA GGTGAGCTTG GAGGACTACA TCAACGACCG CCAGTATGAC	7740
60	TCGCGTGGCC GCTTTGGAGA GCTGCTGCTG CTGCTGCCA CCTTGCAGAG CATCACCTGG	7800
	CAGATGATCG AGCAGATCCA GTTCATCAAG CTCTTCGGCA TGGCCAAGAT TGACAACCTG	7860

	TTGCAGGAGA TGCTGCTGGG AGGTCCGTGC CAAGCCCAGG AGGGGCGGGG TTGGAGTGGG	7920
5	GACTCCCCAG GAGACAGGCC TCACACAGTG AGCTCACCCC TCAGCTCCTT GGCTTCCCCA	7980
	CTGTGCCGCT TTGGGCAAGT TGCTTAACCT GTCTGTGCCT CAGTTTCCTC ACCAGAAAAA	8040
	TGGGAACAAG GCAATGGTCT ATTTGTTAG GCACCGAGAA CCTAGCACGT GCCAGTCACT	8100
10	GTTCTAAGTG CTGGCAATTG AGCAAAGAAC AAGATCTTG CCCTCGGGGA GGCTGTGTGT	8160
	GTGTGAGTAT GTATGGATGC GTGGATATCT GTGTATATGC CCGTATGTGC GTGCATGTGT	8220
15	ATATAAAGCC TCACATTTA TGATTTGAA ATAAACAGGT AATATGAGGG ACACATAGAT	8280
	GCTATAAGTA GGTCAGTTGG CTGCAGCAGA GATGTGGGGG ATGAGGCTGA AAGGTGAGGC	8340
	GGGACCAAAT GGTTGAAGGA CTTGCACTCC AAGGAGCTTT GAGAGCCATT GATTACATCC	8400
20	ATTATGTTAC TATGTGACCA ATACATTACT CATTAGAACAA TTTACGTGAT CTCAGAGCTT	8460
	CCTTATATGC ACCTTGTTC TTTCACTCA CTTTTGTTCT CTTGGTTTTT TGGGGTCCTC	8520
25	TTAACACCCT CATGAAGTCT ATAGATGGGA ATGGTACACC CTAGTTACT AACCCAGGAA	8580
	TAGGTACCCA ACAGGCCTG CCAATATTGG ATGGGCTGGT TGATTGGCCA CGCCTGAGGA	8640
	AGATGGCGTC CCAAGGCCTG AGGTCTGCAT CCCAGACTCT CCATCCTGAT CGACCTTCTC	8700
30	TACCTGCAGG GTCCCCCAGC GATGCACCCC ATGCCACCA CCCCCTGCAC CCTCACCTGA	8760
	TGCAGGAACA TATGGGAACC AACGTCATCG TTGCCAACAC AATGCCCACT CACCTCAGCA	8820
	ACGGACAGAT GTGTGAGTGG CCCCACCCA GGGGACAGGC AGGTGGGCAA ACTCTGGGAT	8880
35	TTTACCTTGC AAAGGGTGAG GATGGGGCTT AAGACAGGAG GCAGGAGAAA GTGGAGTCTA	8940
	GAAGGTAGAA CCAGGATGCA ACAGTTTCT GGGTCCAGG GTAGGGAATA AAGGGCAAGA	9000
40	TTGTCCATTT GTTGAGGCTG TTTATTCACT AAGGTGACTG ACAGCCTTTA CTGAATGAAG	9060
	CCATTGTTGG GATGAGGCAA TCCACTGGAT GAGGTAACCC ATTGGGTGAA GATGTCTTGG	9120
	GTGAGAATTG CATTAGTTGA CATTGTCCAT TAAGTAAAG TGGTCATTGA AGTAAGGCTG	9180
45	CACAGTTGGG TAAGGCTATC CATTAGACAT TAGATGAGAC TACCCATTGG GTCAGGATGT	9240
	CTGCTGGGCT ATTTGGGAGA AGCAGTCCAA GTCTGCATAT CAAATAAATG ATGGAGGAGA	9300
50	TGGGTGGTAG GACCTTCCAG ACCTCATAAA ACTTAGGCTT TATGATCTGG GACTCACAGA	9360
	AGGTTGAGCA ATAAAAGACC TTAGGGATTA TCTGGCTTAA TTAATTCTCT CATTATAG	9420
55	AGGAAGAAAT TAAGTCAAGG TGGGGCAGGG TGGGAGGGGA GAACTTCCC GGGGCTCTTC	9480
	ATTTACTCCC ACAAAAGGCTG GAATTTGAG CAGCCCTGT CTGTCTGTTT GTCCTCCCC	9540
	ACCCCTGAGA CCCCACAGCC CTCACCGCCA GGTGGCTCAG GGTCTGAGCC CTATAAGCTC	9600
60	CTGCCGGGAG CCGTCGCCAC AATCGTCAAG CCCCTCTCTG CCATCCCCCA GCCGACCATC	9660

	ACCAAGCAGG AAGTTATCTA GCAAGCCGCT GGGGCTTGGG GGCTCCACTG GCTCCCCCA	9720
	GCCCCCTAAG AGAGCACCTG GTGATCACGT GGTCACGGCA AAGGAAGACG TGATGCCAGG	9780
5	ACCAGTCCC AAGCAGGAAT GGGAAAGGATG AAGGGCCCGA GAACATGGCC TAAGGCACAT	9840
	CCCACTGCAC CCTGACGCC CGTGCTTGATA ACAAGACTTT GACTTGGGA GACCCTCTAC	9900
10	TGCCCTGGAC AACCTTCTCA TGTTGAAGCC ACTGCCTTCA CCTTCACCTT CATCCATGTC	9960
	CAACCCCCGA CTTCATCCC AAGGACAGCC GCCTGGAGAT GACTTGAGCC TTAC	10014

(2) INFORMATION FOR SEQ ID NO:131:

- 15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 567 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 20 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:131:

25	Met Arg Leu Ser Lys Thr Leu Val Asp Met Asp Met Ala Asp Tyr Ser	
	1 5 10 15	
30	Ala Ala Leu Asp Pro Ala Tyr Thr Thr Leu Glu Phe Glu Asn Val Gln	
	20 25 30	
35	Val Leu Thr Met Gly Asn Gly Pro Ser Ser Pro His Cys Leu Thr Val	
	35 40 45	
40	Ala Leu Leu Gly Ala Trp His Ser Asp Met Met Ile Leu Leu Pro Leu	
	50 55 60	
45	Arg Leu Ala Arg Leu Arg His Pro Leu Arg His His Trp Ser Ile Ser	
	65 70 75 80	
50	Gly Gly Val Asp Ser Ser Pro Gln Gly Asp Thr Ser Pro Ser Glu Gly	
	85 90 95	
55	Thr Asn Leu Asn Ala Pro Asn Ser Leu Gly Val Ser Ala Leu Cys Ala .	
	100 105 110	
60	Ile Cys Gly Asp Arg Ala Thr Gly Lys His Tyr Gly Ala Ser Ser Cys	
	115 120 125	
65	Asp Gly Cys Lys Gly Phe Phe Arg Arg Ser Val Arg Lys Asn His Met	
	130 135 140	
70	Tyr Ser Cys Arg Phe Ser Arg Gln Cys Val Val Asp Lys Asp Lys Arg	
	145 150 155 160	
75	Asn Gln Cys Arg Tyr Cys Arg Leu Lys Lys Cys Phe Arg Ala Gly Met	
	165 170 175	
80	Lys Lys Glu Ala Val Gln Asn Glu Arg Asp Arg Ile Ser Thr Arg Arg	
	180 185 190	
85	Ser Ser Tyr Glu Asp Ser Ser Leu Phe Ser Ile Asn Ala Leu Leu Gln	
	195 200 205	

Ala Glu Val Leu Ser Arg Gln Ile Thr Ser Pro Val Ser Gly Ile Asn
 210 215 220
 5 Gly Asp Ile Arg Ala Lys Lys Ile Ala Ser Ile Ala Asp Val Cys Glu
 225 230 235 240
 Ser Met Lys Glu Gln Leu Leu Val Leu Val Glu Trp Ala Lys Tyr Ile
 245 250 255
 10 Pro Ala Phe Cys Glu Leu Pro Leu Asp Asp Gln Val Ala Leu Leu Arg
 260 265 270
 15 Ala His Ala Gly Glu His Leu Leu Leu Gly Ala Thr Lys Arg Ser Met
 275 280 285
 Val Phe Lys Asp Val Leu Leu Leu Gly Asn Asp Tyr Ile Val Pro Arg
 290 295 300
 20 His Cys Pro Glu Leu Ala Glu Met Ser Arg Val Ser Ile Arg Ile Leu
 305 310 315 320
 25 Asp Glu Leu Val Leu Pro Phe Gln Glu Leu Gln Ile Asp Asp Asn Glu
 325 330 335
 30 Tyr Ala Tyr Leu Lys Ala Ile Ile Phe Phe Asp Pro Asp Ala Lys Gly
 340 345 350
 35 Leu Ser Asp Pro Gly Lys Ile Lys Arg Leu Arg Ser Gln Val Gln Val
 355 360 365
 40 Ser Leu Glu Asp Tyr Ile Asn Asp Arg Gln Tyr Asp Ser Arg Gly Arg
 370 375 380
 45 Phe Gly Glu Leu Leu Leu Leu Leu Pro Thr Leu Glu Ser Ile Thr Trp
 385 390 395 400
 50 Gln Met Ile Glu Gln Ile Gln Phe Ile Lys Leu Phe Gly Met Ala Lys
 405 410 415
 55 Ile Asp Asn Leu Leu Gln Glu Met Leu Leu Gly Gly Pro Cys Gln
 420 425 430
 60 Ala Gln Glu Gly Arg Gly Trp Ser Gly Asp Ser Pro Gly Asp Arg Pro
 435 440 445
 His Thr Val Ser Ser Pro Leu Ser Ser Leu Ala Ser Pro Leu Cys Arg
 450 455 460
 65 Phe Gly Gln Val Ala Gly Ser Pro Ser Asp Ala Pro His Ala His His
 465 470 475 480
 Pro Leu His Pro His Leu Met Gln Glu His Met Gly Thr Asn Val Ile
 485 490 495
 70 Val Ala Asn Thr Met Pro Thr His Leu Ser Asn Gly Gln Met Cys Glu
 500 505 510
 75 Trp Pro Arg Pro Arg Gly Gln Ala Ala Thr Pro Glu Thr Pro Gln Pro
 515 520 525

Ser Pro Pro Gly Gly Ser Gly Ser Glu Pro Tyr Lys Leu Leu Pro Gly
 530 535 540
 Ala Val Ala Thr Ile Val Lys Pro Leu Ser Ala Ile Pro Gln Pro Thr
 545 550 555 560
 Ile Thr Lys Gln Glu Val Ile
 565

10

(2) INFORMATION FOR SEQ ID NO:132:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 470 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:132:

AAGTAAGCCT	TGTTTTCCA	CACTCATTCT	CCCAGGTTT	CTTGATAG	GCTTACTTTT	60
CCATGCTGGA	GGAGGGCTA	TCCCTTCATT	TTGCCTCTCC	CGCTTCCCTC	CCTCTCCCCC	120
TCCCCCTGCT	TTCTCTCCCT	CTGCACTTTG	TGAACTGCTG	CTGCAGTGCT	GAAGTCCAAA	180
GTTCAGTAAC	TTGCTAACGCA	CACAGATAAA	TATGAACCTT	GGAGAATTAA	CCAATGTAAA	240
CAGATAGCCA	AGGGTCCCTT	TATCAGCACT	GGCTCAGGAC	AGTCGTGGGG	GGTCTGAAGT	300
GGCTCAATT	TGTATTTGT	TTTTTTGGG	GGGGTGTAAA	GGCGGGAGGC	TGCGCTGTGC	360
CCGCTGCTGA	CAGTCGGCG	TGTTACCTCG	GGAACATGGT	GTAGGGAAGC	TGGAAGCAGG	420
ATAACGTGGA	ACTCAACCCA	AGAAACGCCA	GCCTGAAGAC	CATGGTCTCG		470

(2) INFORMATION FOR SEQ ID NO:133:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 467 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:133:

TCACAGCTAT	TAGCTCATCG	CTGCCAAATT	GCCCCTTAC	CTAGGCTTGT	GTCACTTCA	60
CCTTCTCATT	CTCTTACTTT	TACATTCTTC	CTTGATATTT	TGCTTTTCA	ACTTTGGAA	120
ATTTCTTTCT	CTCTTCTACC	CCTCCTCATA	TTCCCTCTGCA	CTCCCCCCTC	TCTAACTCAT	180
GCACTTTGTG	GGGTCAAAG	TTCAGTAACT	TGCAAAGCAC	AGGGATAAAAG	ATGAACCTTG	240
GAAGATTTAC	TCTGCTCTGA	TGTAAACAGA	GAGTGACAAG	GGTCCCTTAT	CTATGTCTCA	300
GAGAAGCCTG	TCCGGGGGT	GACCACTTGC	TGGTTGTGGC	TGCACAGTGT	GTTTTTTGG	360
GGGGGAGGAG	GAAACAGAAG	GTGGGTAGAG	CATGGACTCC	CGCCCGCTGA	TCCGTGTTAC	420

AGCCGCAGAT GGTGAGGCAG TAGAAGGCAA CAGACAGGAT GGCGTCT

467

5 (2) INFORMATION FOR SEQ ID NO:134:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 479 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
10 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:134:

15	TTTCGGGGGT GGGACCCAAC GCTGCTCTCC TGATGGCCTC CCTGGCTCCC AGCACCTTCC	60
20	ATCCCAGCTG CTCAGGGCCC CTCACCTGCG CCTCCCCCAC CCTCCCTCT GCCCACTCCC	120
25	ATCGCAGGCC ATAGCTCCCT GTCCCTCTCC GCTGCCATGA GGCCCTGCACT TTGCAGGGCT	180
30	GAAGTCCAAA GTTCAGTCCC TTGCTAAGC ACACGGATAA ATATGAACCT TGGAGAATT	240
35	CCCCAGCTCC AATGTAAACA GAACAGGCAG GGGCCCTGAT TCACGGGCCG CTGGGGCCAG	300
40	GGTTGGGGGT TGGGGGTGCC CACAGGGCTT GGCTAGTGGG GTTTGGGGG GGCAGTGGGT	360
45	GCAAGGAGTT TGGTTTGTT CTGCCGGCCG GCAGGCAAAC GCAACCCACG CGGTGGGGA	420
50	GGCGGCTAGC GTGGTGGACC CGGGCCGCGT GGCCCTGTGG CAGCCGAGCC ATGGTTTCT	479

(2) INFORMATION FOR SEQ ID NO:135:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 605 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:135:

40	TGGGGCCTGG GATTAGGTT TCTAAATCGT GGGCCATGGG GCAGCCTTAT CTCTGCAAAA	60
45	GCATTGAGGG TAGAAGTCAA TGATTGGGA AGTTATTGAA TTAGGGATC TCGGAGGTAG	120
50	GCTGTCAGTG CCTGATAGTA TCAGTTAGAA TGCCTGACTT GGGGTGACAA TGGCTTGGAG	180
55	GGGTGGGTGA GTCAAGGGTC AAATGAGTGC CCGTGAGTC TGATGCCTGC CTTGTACAAT	240
60	TGATAACTGA ACATCGGTGA GTAGGGCCC CAGCAGTTGT ATTAGCACC CCGGGTGTCA	300
	GCCAGAAACC AACAAACAGC CAAATCCCTG CAGCCCCGCC CAGCCTATCC ACCGGCGGGG	360
	GACCGATTAA CCATTAACCC CCACCCCTCC CCGGCAGAGC CTCCACCCCT TCACAGAGGC	420
	TAGGCCAAGA CTCCCAGCAG ATCTTCCCAG AGGACGGTTT GAAAGGAAGG CAGAGAGGGC	480
	ACTGGGAGGA GGCAGTGGGA GGGCGGAGGG CGGGGGCCTT CGGGGTGGC GCCCAGGGTA	540
	GGGCAGGTGG CCGCGGCGTG GAGGCAGGGA GAATGCGACT CTCCAAAACC CTCGTCGACG	600
	ACATG	605

(2) INFORMATION FOR SEQ ID NO:136:

- 5 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 478 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:136:

TCCTGGAGAG	TGGGACCCAG	CGCCGCACCC	AGAGGCCTCC	TGGCTCCTGC	TGCCTCTAGC	60
CCTGCGCCCC	TGGCCCCTCT	CCACCTCCCC	CACCCTCCCT	TCTGCTCACT	CCCAATTGCA	120
GGCCATGACT	CCGGTCCGCG	TCCCTCTCAC	CCCCATGAGG	CCTGCACTTG	CAAGGCTGAA	180
GTCCAAGTT	CAGTCCCTTC	GCTAAGCGCA	CGGATAAAATA	TGAACCTTGG	AGAATTCCC	240
CAGCTCCAAT	GTAAACAGAG	CAGGCAGGGG	CCCTGATTCA	CTGGCCGCTG	GGGCCAGGGT	300
TGGGGGCTGG	GGGTGCCAC	AGAGCTTGAC	TAGTGGGATT	TGGGGGGGCA	GTGGGTGCAG	360
CGAGCCCGGT	CCGTTGACTG	CCAGCCTGCC	GGCAGGTAGA	CACCGGCCGT	GGGTGGGGGA	420
GGCGGCTAGC	TCAGTGGCCT	TGGGCCGCGT	GGCTGGTGGC	AGCGGAGCCA	TGGTTTCT	478

30 (2) INFORMATION FOR SEQ ID NO:137:

- 35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 622 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:137:

TGGGCTTGGG	TGTTAGGTTT	CCAGTTCAAG	CGACCCAGGA	CAGCTTTATC	TCAAATTGAG	60
GATAGAACGTC	AATGATCTGG	GACGTGATTG	GCTTAGGGCT	TCATAGTGGT	AGGCTTGCCA	120
GTGTCTAAC	ATGTCAGCTG	GGTTGTCCAC	CTTGGTGAGA	CTTGGGGGCT	GCTGAGGCAA	180
GGGGTCCAAC	CAATGCCAGT	CCTGTTGGGT	GCCTGCCTTG	GAAGATTGGT	AAGTGACTAT	240
TAATGAGCGG	GAGGTGGGGG	GGGGCAACA	GTTGTAATTA	GCACCCCAGG	TGTCAGTCAG	300
AAACCAACAA	ACAGCAAAT	CCTCGTGGCT	CCACCCAGCC	TACCCAGCAA	CGGGGGTGAT	360
TAACCATTAA	CTCCTACCCC	TCCCCACAGA	GCCTCCACCC	TCTGCAGAGG	CTAGGCCAGG	420
ACGCCAGGCT	GAGTCTCCA	GAGGACAGTT	TGAAAGAGAG	GAAGGCAGAG	AAGGGACCTG	480
GGAGGAGGCA	GGAGGAGGGC	GGGGACGGGG	GGGGCTGGGG	CTCAGCCAG	GGGCTTGGGT	540
GGCATCCTGG	GCCGGGCAGG	ACAGGGGGCT	AAGGCGTGGG	TAGGGGAGAA	TGCGACTCTC	600
TAAAACCCCTT	GCCGGCGATA	TG				622

5 (2) INFORMATION FOR SEQ ID NO:138:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 470 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:138:

TCTTGGGCAG TGGGACCAGC GCTGCTCCC AAGGCCTCCT GGCTCCTGGT GCCTCTCTCC 60
CTGCGCCCC CT GTTCCCGCT CCACCTCCCC CACCCGCCCT TCTGCTCACT CCCAATTGCA 120
15 AGCCATGGCT CCCGGTCCGG TCCCTCTCGC TGCTGTGAGG CCTGCACTTG CAAGGCTGAA 180
GTCCAAGTT CAGTCCCTTC GCTAAGCACA CGGATAAAATA TGAACCTTGG AGAATTCCC 240
20 CAGCTCCAAT GTAAACAGAG CAGCAGGGGG CCCTGATTCA CTAGCCGCTG GGGCCAGGGT 300
TGGGGGTTGG GGGTGCCAC AGGGCTTGAC TAGTGGGATT TGGGGGAGCA GTGGGTGCAG 360
25 CGAGCCTGGT CCGTTGACTG CCAGCAGTAG ACACCGGCCG TGTGTGGGGG AGGCAGCTAG 420
30 CTCAGTGGCC TTGGGCCGCG TGGCCTGGCG GTAGAGGAGC CATGGTTCT 470

35 (2) INFORMATION FOR SEQ ID NO:139:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 557 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:139:

Met Val Ser Lys Leu Thr Ser Leu Gln Gln Glu Leu Leu Ser Ala Leu
1 5 10 15
Leu Ser Ser Gly Val Thr Lys Glu Val Leu Val Gln Ala Leu Glu Glu .
20 25 30
45 Leu Leu Pro Ser Pro Asn Phe Gly Val Lys Leu Glu Thr Leu Pro Leu
35 40 45
Ser Pro Gly Ser Gly Ala Glu Pro Asp Thr Lys Pro Val Phe His Thr
50 55 60
50 Leu Thr Asn Gly His Ala Lys Gly Arg Leu Ser Gly Asp Glu Gly Ser
65 70 75 80
55 Glu Asp Gly Asp Asp Tyr Asp Thr Pro Pro Ile Leu Lys Glu Leu Gln
85 90 95
Ala Leu Asn Thr Glu Glu Ala Ala Glu Gln Arg Ala Glu Val Asp Arg
100 105 110
60 Met Leu Ser Glu Asp Pro Trp Arg Ala Ala Lys Met Ile Lys Gly Tyr
115 120 125

	Met Gln Gln His Asn Ile Pro Gln Arg Glu Val Val Asp Val Thr Gly			
	130	135	140	
5	Leu Asn Gln Ser His Leu Ser Gln His Leu Asn Lys Gly Thr Pro Met			
	145	150	155	160
	Lys Thr Gln Lys Arg Ala Ala Leu Tyr Thr Trp Tyr Val Arg Lys Gln			
10	165	170	175	
	Arg Glu Ile Leu Arg Gln Phe Asn Gln Thr Val Gln Ser Ser Gly Asn			
	180	185	190	
15	Met Thr Asp Lys Ser Ser Gln Asp Gln Leu Leu Phe Leu Phe Pro Glu			
	195	200	205	
	Phe Ser Gln Gln Ser His Gly Pro Gly Gln Ser Asp Asp Ala Cys Ser			
	210	215	220	
20	Glu Pro Thr Asn Lys Lys Met Arg Arg Asn Arg Phe Lys Trp Gly Pro			
	225	230	235	240
	Ala Ser Gln Gln Ile Leu Tyr Gln Ala Tyr Asp Arg Gln Lys Asn Pro			
25	245	250	255	
	Ser Lys Glu Glu Arg Glu Ala Leu Val Glu Glu Cys Asn Arg Ala Glu			
	260	265	270	
30	Cys Leu Gln Arg Gly Val Ser Pro Ser Lys Ala His Gly Leu Gly Ser			
	275	280	285	
	Asn Leu Val Thr Glu Val Arg Val Tyr Asn Trp Phe Ala Asn Arg Arg			
	290	295	300	
35	Lys Glu Glu Ala Phe Arg Gln Lys Leu Ala Met Asp Ala Tyr Ser Ser			
	305	310	315	320
	Asn Gln Thr His Ser Leu Asn Pro Leu Leu Ser His Gly Ser Pro His			
40	325	330	335	
	His Gln Pro Ser Ser Pro Pro Asn Lys Leu Ser Gly Val Arg Tyr			
	340	345	350	
45	Ser Gln Gln Gly Asn Asn Glu Ile Thr Ser Ser Ser Thr Ile Ser His			
	355	360	365	
	His Gly Asn Ser Ala Met Val Thr Ser Gln Ser Val Leu Gln Gln Val			
	370	375	380	
50	Ser Pro Ala Ser Leu Asp Pro Gly His Asn Leu Leu Ser Pro Asp Gly			
	385	390	395	400
	Lys Met Ile Ser Val Ser Gly Gly Leu Pro Pro Val Ser Thr Leu			
55	405	410	415	
	Thr Asn Ile His Ser Leu Ser His His Asn Pro Gln Gln Ser Gln Asn			
	420	425	430	
60	Leu Ile Met Thr Pro Leu Ser Gly Val Met Ala Ile Ala Gln Ser Leu			
	435	440	445	

Asn Thr Ser Gln Ala Gln Ser Val Pro Val Ile Asn Ser Val Ala Gly
 450 455 460
 5 Ser Leu Ala Ala Leu Gln Pro Val Gln Phe Ser Gln Gln Leu His Ser
 465 470 475 480
 Pro His Gln Gln Pro Leu Met Gln Gln Ser Pro Gly Ser His Met Ala
 485 490 495
 10 Gln Gln Pro Phe Met Ala Ala Val Thr Gln Leu Gln Asn Ser His Met
 500 505 510
 Tyr Ala His Lys Gln Glu Pro Pro Gln Tyr Ser His Thr Ser Arg Phe
 515 520 525
 15 Pro Ser Ala Met Val Val Thr Asp Thr Ser Ser Ile Ser Thr Leu Thr
 530 535 540
 20 Asn Met Ser Ser Ser Lys Gln Cys Pro Leu Gln Ala Trp
 545 550 555

2 (2) INFORMATION FOR SEQ ID NO:140:

- 25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 516 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:140:

Met Asp Met Ala Asp Tyr Ser Ala Ala Leu Asp Pro Ala Tyr Thr Thr
 1 5 10 15
 35 Leu Glu Phe Glu Asn Val Gln Val Leu Thr Met Gly Asn Gly Pro Ser
 20 25 30
 40 Ser Pro His Cys Leu Thr Val Ala Leu Leu Gly Ala Trp His Ser Asp
 35 40 45
 Met Met Ile Leu Leu Pro Leu Arg Leu Ala Arg Leu Arg His Pro Leu ,
 50 55 60
 45 Arg His His Trp Ser Ile Ser Gly Gly Val Asp Ser Ser Pro Gln Gly
 65 70 75 80
 50 Asp Thr Ser Pro Ser Glu Gly Thr Asn Leu Asn Ala Pro Asn Ser Leu
 85 90 95
 55 Gly Val Ser Ala Leu Cys Ala Ile Cys Gly Asp Arg Ala Thr Gly Lys
 100 105 110
 His Tyr Gly Ala Ser Ser Cys Asp Gly Cys Lys Gly Phe Phe Arg Arg
 115 120 125
 60 Ser Val Arg Lys Asn His Met Tyr Ser Cys Arg Phe Ser Arg Gln Cys
 130 135 140
 Val Val Asp Lys Asp Lys Arg Asn Gln Cys Arg Tyr Cys Arg Leu Lys
 145 150 155 160

	Lys Cys Phe Arg Ala Gly Met Lys Lys Glu Ala Val Gln Asn Glu Arg			
	165	170	175	
5	Asp Arg Ile Ser Thr Arg Arg Ser Ser Tyr Glu Asp Ser Ser Leu Phe			
	180	185	190	
	Ser Ile Asn Ala Leu Leu Gln Ala Glu Val Leu Ser Arg Gln Ile Thr			
10	195	200	205	
	Ser Pro Val Ser Gly Ile Asn Gly Asp Ile Arg Ala Lys Lys Ile Ala			
	210	215	220	
15	Ser Ile Ala Asp Val Cys Glu Ser Met Lys Glu Gln Leu Leu Val Leu			
	225	230	235	240
	Val Glu Trp Ala Lys Tyr Ile Pro Ala Phe Cys Glu Leu Pro Leu Asp			
	245	250	255	
20	Asp Gln Val Ala Leu Leu Arg Ala His Ala Gly Glu His Leu Leu Leu			
	260	265	270	
25	Gly Ala Thr Lys Arg Ser Met Val Phe Lys Asp Val Leu Leu Leu Gly			
	275	280	285	
	Asn Asp Tyr Ile Val Pro Arg His Cys Pro Glu Leu Ala Glu Met Ser			
	290	295	300	
30	Arg Val Ser Ile Arg Ile Leu Asp Glu Leu Val Leu Pro Phe Gln Glu			
	305	310	315	320
	Leu Gln Ile Asp Asp Asn Glu Tyr Ala Tyr Leu Lys Ala Ile Ile Phe			
	325	330	335	
35	Phe Asp Pro Asp Ala Lys Gly Leu Ser Asp Pro Gly Lys Ile Lys Arg			
	340	345	350	
	Leu Arg Ser Gln Val Gln Val Ser Leu Glu Asp Tyr Ile Asn Asp Arg			
	355	360	365	
40	Gln Tyr Asp Ser Arg Gly Arg Phe Gly Glu Leu Leu Leu Leu Leu Pro			
	370	375	380	
45	Thr Leu Glu Ser Ile Thr Trp Gln Met Ile Glu Gln Ile Gln Phe Ile			
	385	390	395	400
	Lys Leu Phe Gly Met Ala Lys Ile Asp Asn Leu Leu Gln Glu Met Leu			
	405	410	415	
50	Leu Gly Gly Ser Pro Ser Asp Ala Pro His Ala His His Pro Leu His			
	420	425	430	
	Pro His Leu Met Gln Glu His Met Gly Thr Asn Val Ile Val Ala Asn			
55	435	440	445	
	Thr Met Pro Thr His Leu Ser Asn Gly Gln Met Cys Glu Trp Pro Arg			
	450	455	460	
60	Pro Arg Gly Gln Ala Ala Thr Pro Glu Thr Pro Gln Pro Ser Pro Pro			
	465	470	475	480

Gly Gly Ser Gly Ser Glu Pro Tyr Lys Leu Leu Pro Gly Ala Val Ala
485 490 495

5 Thr Ile Val Lys Pro Leu Ser Ala Ile Pro Gln Pro Thr Ile Thr Lys
500 505 510

Gln Glu Val Ile
515

10

(2) INFORMATION FOR SEQ ID NO:141:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:141:

20

GCGGGACCGG ATCAGCA

17

25

(2) INFORMATION FOR SEQ ID NO:142:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:142:

30

Arg Asp Arg Ile Ser
1 5

35

(2) INFORMATION FOR SEQ ID NO:143:

40

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:143:

50

GCGGGACTGG ATCAGCA

17

55

(2) INFORMATION FOR SEQ ID NO:144:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:144:

Ala Glu Val Leu Ser Arg Gln
1 5

(2) INFORMATION FOR SEQ ID NO:145:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: 16
- (D) OTHER INFORMATION: /note= "N = C or T"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:145:

GCGGAGGTCC TGTCCNGACA GGTACCGGGG

30

(2) INFORMATION FOR SEQ ID NO:146:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: 8
- (D) OTHER INFORMATION: /note= "N = C or T"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:146:

AAAGCAANGA GAGAT

15

(2) INFORMATION FOR SEQ ID NO:147:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 3
- (D) OTHER INFORMATION: /note= "X = R or any amino acid"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:147:

Lys Gln Xaa Glu
1